Ultra-Dense Networks in 5G and Beyond: Challenges and Promising Solutions

Mohammed Elbayoumi

A Thesis

in

The Concordia Institute

for

Information Systems Engineering

Presented in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy (Information and Systems Engineering) at
Concordia University
Montréal, Québec, Canada

March 2023

Concordia University

School of Graduate Studies

This is to certify that the thesis prepared

By: **Mohammed Mohammed Elbayoumi Elsayed**

Entitled: **Ultra-Dense Networks in 5G and Beyond: Challenges and Promising Solutions**

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Information and Systems Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

_____ Chair
Dr. Anjan Bhowmick

_____ External Examiner
Dr. Cheng Li

_____ External to Program
Dr. Yousef Shayan

_____ Examiner
Dr. Chadi Assi

_____ Examiner
Dr. Jamal Bentahar

_____ Supervisor
Dr. Walaa Hamouda

_____ Supervisor
Dr. Amr Youssef

Approved by _____
Dr. Zachary Patterson, Graduate Program Director

March 7^{th}, 2023
Date of Defence

_____
Dr. Mourad Debbabi, Dean
Gina Cody School of Engineering and Computer Science

Abstract

Ultra-Dense Networks in 5G and Beyond: Challenges and Promising Solutions

Mohammed Elbayoumi, Ph.D.

Concordia University, 2023

Ultra-Dense Network (UDN) is one of the promising and leading directions in Fifth Generation and beyond (5GB) networks. In UDNs, Small Cells (SCs) or Small Base Stations (SBSs) such as microcells, picocells, or femtocells are deployed in high densities where inter-site distances are within the range of few or tens of meters. UDNs also require that SCs are typically deployed in relatively large densities compared to the Human-Type Communication Users (HTCUs) such as smartphones, tablets, and/or laptops. Such SCs are characterized by their low transmission powers, small coverage areas, and low cost. Hence, the deployment of the SCs can be done either by the cellular network operators or by the customers themselves within their premises to maintain certain levels of Quality of Service (QoS). However, the randomness of the deployment of the SCs along with the small inter-site distances may degrade the achievable performance due to the uncontrolled Inter-Cell Interference (ICI). Therefore, idle mode capability is an inevitable feature in the high-density regime of SCs. In idle mode, a SC is switched off to prevent ICI when no user is associated to it. In doing so, we can imagine the UDN as a mobile network that keeps following the users to remain as close as possible to them.

In 5G, different use cases are required to be supported such as enhanced Mobile Broad-Band (eMBB), Ultra-Reliable and Low-Latency Communication (URLLC), and massive Machine-Type Communication (mMTC). On one hand, the inevitable upcoming era of smart living requires unprecedented advances in enabling technologies to support the main building blocks of this era which are Internet of Things (IoT) devices. Machine-Type Communication (MTC), the cellular version of Machine-to-Machine (M2M) communication, constitutes the main enabling technology to support communications among such devices with minimal or even without human intervention. The massive number of these devices, Machine-Type Communication Devices (MTCDs), and the immense amount of traffic generated by them require a paramount shift in cellular and non-cellular wireless technologies to achieve the required connectivity. On the other hand, the sky-rocketing number of data hungry applications installed on human-held devices, or HTCUs, such as video conferencing and virtual reality applications require their own advances in the wireless infrastructure in terms of high capacity, enhanced reliability, and reduced latency.

Throughout this thesis, we exploit the UDN infrastructure integrated with other 5G resources and enabling technologies to explore the possible opportunities in supporting both HTC and MTC, either solely or simultaneously. Given the shorter distances between transmitters and receivers encountered in UDNs, more realistic models of the path loss must be adopted such as the Stretched Exponential Path Loss (SEPL) model. We use tools from stochastic geometry to formulate novel mathematical frameworks that can be used to investigate the achievable performance without having to rely on extensive time-consuming Monte-Carlo simulations. Besides, the derived analytical expressions can be used to tune some system parameters or to propose some approaches/techniques that can be followed to optimize the performance of the system under certain circumstances.

Tackling practical scenarios, the complexity, or sometimes in-feasibility, of providing unlimited backhaul capacity for the massive number of SCs must be considered. In this regard, we adopt multiple-association where each HTCU is allowed to associate with multiple SCs. By doing so, we carefully split the targeted traffic among several backhaul links to mitigate the bottleneck forced by limited backhaul capacities. It is noteworthy that for coexisting MTCDs with the HTCUs, activating more SCs would allow more

MTCDs to be supported without introducing additional ICI towards the HTCUs.

Targeting different application, multiple-association can be also adopted to tackle computation-intensive applications of HTCUs. In particular, for applications such as augmented reality and environment recognition that require heavy computations, a task is split and partially offloaded to multiple SCs with integrated Edge Computing Servers (ECSs). Then, the task partitions are processed in parallel to reduce the end-to-end processing delay. Based on relative densities between HTCUs and SCs, we use tools from stochastic geometry to develop an offline adaptive task division technique that further reduces the average end-to-end processing delay per user.

With the frequent serious data breaches experienced in recent years, securing data has become more of a business risk rather than an information technology (IT) issue. Hence, we exploit the dense number of SCs found in UDN along with Physical Layer Security (PLS) protocols to secure data transfer. In particular, we again adopt multiple-association and split the data of HTCUs into multiple streams originating from different SCs to prevent illegitimate receivers from eavesdropping.

To support massive number of MTCDs, we deploy the Non-Orthogonal Multiple-Access (NOMA) technique. Using power NOMA, more than one device can be supported over the same frequency/time resource and their signals are distinguished at the receiver using Successive Interference Cancellation (SIC). In the same scope, exploiting the available resources in 5G and beyond networks, we investigate a mMTC scenario in an UDN operating in the Millimeter Wave (mmWave) band and supported by wireless backhauling. In doing so, we shed lights on the possible gains of utilizing the mmWave band where the severe penetration losses of mmWave can be exploited to mitigate the significant ICI in UDNs. Also, the vast bandwidth available in the mmWave band helps to allocate more Resource Blocks (RBs) per SCs which corresponds to supporting more MTCDs.

Acknowledgments

First and foremost, all praises and thanks are due to Allah (God), the Almighty, the Most Gracious, and the Most Merciful, for His countless blessings, and prayers and peace be upon Muhammad His servant and messenger. I could have never finished this thesis without the faith I have in Almighty Allah and His prophet Muhammad (PBUH).

I would like first to express my sincere gratitude to my supervisors, Professor Amr Youssef and Professor Walaa Hamouda. I will always be indebted for your continuous support, encouragement, patience, guidance, and immense knowledge. It is my honor to be supervised by such great professors and mentors like you. There are not enough words to express my gratitude for your personal and technical support and advice through my PhD journey. Dear professors, I can just say that the greatest reward of my PhD is getting to know you, thank you Dr. Hamouda and Dr. Youssef.

Also, I would like to sincerely thank the examining committee: Dr. Chadi Assi, Dr. Jamal Bentahar, and Dr. Yousef Shayan for their helpful insights, valuable discussion, and guidance throughout each milestone of the PhD journey. I am also thankful to Concordia Graduate Student Support Program (GSSP) for funding my PhD's studies.

To my friends, Salah Elhoushy and Mohammed Ibrahim, thank you for your support in the most stressful moments and for the pleasant times and fruitful discussions we had over the past four years. Wishing you all happiness and success throughout your life.

No words can express my gratitude to my beloved mother Samia Awad, thank you for your unlimited and unconditional support and love. To my mother and the soul of my beloved father, to my siblings, and to my nephews and nieces; you have always been loving, caring, and supportive throughout my entire life, I love you all.

Last, but not least, there is a special person in my life, my beloved wife Sally Shawky, to whom I owe a lot for her love and encouragement. Without your great sacrifices, I would not have reached this level in my research career. Thank you for always being there for me, for bringing your beloved family as a second family into my life, and for bringing our lovely and beautiful children Yussuf and Farida. My dear little family, you mean everything to me.

To my Beloved Mother,

To my Soulmate, Sally Shawky, and

To my inspiring kids, Yussuf and Farida.

Table of Contents

List of Figures

List of Tables

List of Acronyms

3GPP	Third Generation Partnership Project
5G	Fifth Generation
5GB	5G and Beyond
AL	Access Link
ASE	Area Spectral Efficiency
AWGN	Additive White Gaussian Noise
BH	Backhaul Link
BS	Base Station
C2A	Connect-to-Active
C2C	Connect-to-closest
CAGR	Compound Annual Growth Rate
CCDF	Complementary Cumulative Distribution Function
CDF	Cumulative Distribution Function
CPU	Central Processing Unit
CRAN	Cloud Radio Access Network
CSI	Channel State Information
ECS	Edge Computing Server
EESM	Effective Exponential SNR Mapping
eMBB	Enhanced Mobile Broad-Band
ETSI	European Telecommunications Standards Institute

FDMA	Frequency Division Multiple Access
HPPP	Homogeneous Poisson Point Process
HTC	Human-Type Communication
HTCU	Human-Type Communication User
i.i.d.	Independently and Identically Distributed
IAB	Integrated Access and Backhaul
ICI	Inter-Cell Interference
IMT	International Mobile Telecommunications
INI	Inter-NOMA Interference
IoT	Internet of Things
IT	Information Technology
LoS	Line-of-Sight
M2M	Machine-to-Machine
MEC	Multi-Access Edge Computing
MIMO	Multiple-Input Multiple-Output
mMIMO	massive Multiple-Input Multiple-Output
mMTC	massive Machine-Type Communication
mmWave	Millimeter-Wave
MTC	Machine-Type Communication
MTCD	Machine-Type Communication Device
NLoS	Non Line-of-Sight
NOMA	Non-Orthogonal Multiple-Access
NR	New Radio
OMA	Orthogonal Multiple-Access
PCP	Poisson Cluster Process
pdf	probability density function
PD-NOMA	Power-Domain NOMA

PGFL	Probability Generating Functional
PLS	Physical Layer Security
pmf	probability mass function
PPP	Poisson Point Process
QoS	quality of service
RB	Resource Block
RRH	Remote Radio Head
SC	Small Cell
SCs	Small Cells
SEPL	Stretched Exponential Path Loss
SIC	Successive Interference Cancellation
SINR	Signal-to-Interference plus Noise Ratio
SIR	Signal-to-Interference Ratio
SNR	Signal-to-Noise Ratio
SWIPT	Simultaneous Wireless Information and Power Transfer
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
TTI	Transmit Time Interval
UDN	Ultra-Dense Network
UE	User Equipment
URLLC	Ultra-Reliable and Low-Latency Communications

List of Common Symbols

Φ_s	PPP of SCs
λ_s	Density of SCs
Φ_h	PPP of HTCUs
λ_h	Density of HTCUs
Φ_m	PPP of MTCDs
λ_m	Density of MTCDs
λ_b	Density of blockages
λ_e	Density of eavesdroppers
ρ_m	Activity Ratio of MTCDs
r_k	Distance to the kth closest SC
h	Channel power gain of small scale fading
S	Useful Signal
I	Sensed ICI
σ^2	Thermal noise power
γ	Achievable SINR
$\mathbb{E}[.]$	Mathematical Expectation
$\mathbb{P}[.]$	Probability of event
$\mathcal{L}\{.\}$	Laplace transform
$\Gamma[.]$	Gamma function
$f_X(.)$	Probability density (mass) function of random variable X

$F_X(.)$	Cumulative distribution function of random variable X
N_{RB}	Number of resource blocks
M	*MultiCell* size
ρ	Limited backhaul capacity
λ_m^a	Density of active MTCDs
λ_s^a	Density of active SCs
p_k^a	probability of cell activation on *tier-k*
p_k^{all}	probability of cell activation on all tiers from *tier-1* up to *tier-k*
λ_k^a	Density of active SCs on *tier-k*
λ_k^{all}	Density of active SCs on all tiers from *tier-1* to *tier-k*
α	SEPL tunable parameter
β	SEPL tunable parameter
ω	Fraction of bandwidth for HTC
N	Number of associated users/devices per SC
λ	Average number of users per SC
N_m^{cell}	Number of associated MTCDs with an SC
$N_{m,s}^{cell}$	Average number of supported MTCDs per SC
λ_m^s	Density of supported MTCDs
R	Instantaneous rate per SC per Hz on the radio link
$\hat{R}$	Instantaneous rate per SC per Hz under limited backhaul capacity
$\bar{R_k^M}$	Average achievable rate per SC on *tier-k* (with *MultiCell* size M)
$\hat{R_k^M}$	Average achievable rate per SC on *tier-k* under limited backhaul
$\mathcal{T}_h^d$	Average downlink ASE of HTC
$\mathcal{T}_h^u$	Average uplink ASE of HTC
$\bar{R}_m$	Average rate per MTCD
$\mathcal{T}_m^u$	Average uplink ASE of MTC
W	Available bandwidth for task offloading

γ_{th}	Offloading link activation-SINR-threshold
η	Power-law path loss exponent
p_k	Probability of *tier-k* link association
$p_k^{activation}$	Probability of *tier-k* link activation
B_u	Task transmit-size in Bits
C_u	Task computation size in CPU cycles
f_{loc}	HTCU local processing capacity
t_{loc}	Local processing time of the whole task C_u
f_{ecs}	ECS processing capacity
t_{ecs}	ECS processing time of the whole task C_u
α_k^u	Fraction of task to be offloaded through *tier-k*
$t_{k,tran}^u$	Offloading transmission time of task fraction through *tier-k*
R_k^u	Uplink transmission rate on *tier-k*
N_t	Number of connected tiers
x_i^M	Specific state combination of different tier activation
$\mathcal{M}_0^{x_i^M}$	Set of non-associated tiers
$\mathcal{M}_1^{x_i^M}$	Set of associated and active tiers
$\mathcal{M}_2^{x_i^M}$	Set of associated but inactive tiers
$\mathbb{P}_{idle}^j$	Idle mode probability on *tier-j*
$\mathbb{P}_a^j$	Probability of SC activation on *tier-j*
$\mathcal{R}_s^j$	Secrecy rate on *tier-j*
$\mathcal{R}_m^j$	Rate on the main link of *tier-j*
$\mathcal{R}_{e^*}^j$	Leakage rate from *tier-j* to the MDE
R_i^O	Rate per user i per one Hz adopting OMA
R_k^N	Rate per user k in the NOMA cluster per one Hz
ζ	SIC error propagation factor
ε_k	Power fraction allocated to user k

$I_{k,h}^{N}$	INI on user k due to messages of higher allocated powers
R_{cell}	Rate per SC
c_i	Normalized SIR of user i
T	SIR threshold
N_1	Number of group-1 users per SC
N_2	Number of group-2 users per SC
Z	Number of NOMA pairs per SC
p_a	Probability of SC activation
N_s	Number of transmitting antennas per SC
N_c	Number of transmitting antennas per CPU
$\|h^{AL}\|^2$	Combined channel and antenna power gain on the AL
$\mathcal{A}_q$	Average combined channel and antenna power gain on the AL
θ_{alg}	Beamwidth of aligned beam
$\bar{L}$	Average blockage length
P_{LoS}	Probability of LoS
$\mathcal{B}_b$	Blockage parameter
α_z^{AL}	Path loss exponent on AL
β_z^{AL}	Path loss intercept on AL
ζ_z^{AL}	Path loss between SC and MTCD
$\boldsymbol{g}_{c_s s}^{BH}$	the channel vector between SC and its serving CPU
p_m^{out}	Outage probability of MTCD
p_m^{assoc}	Probability of association of an MTCD with an SC
p_s^{out}	Probability of SC with no associated MTCDs
p_s^{idle}	Idle mode probability of an SC
M_m	Number of supported MTCDs per SC
$\bar{M}_m$	Average number of supported MTCDs per SC
τ	Achievable rate per MTCD

γ'	SINR activation threshold
τ'	Achievable rate per active MTCD
W_{RB}^{AL}	available AL bandwidth per R
C_s^{AL}	Achievable capacity per SC on the AL
C_s^{BH}	Achievable capacity per SC on the BH
C_s^{bound}	Bounded achievable capacity per SC

Chapter 1

Introduction

1.1 Overview

In order to achieve the smart living of the future, it is necessary for all objects to be connected to the Internet. This infrastructure is called the Internet of Things (IoT), and it is made up of devices such as sensors and actuators, known as IoT devices. The main technology that enables these devices to communicate with each other is called Machine-Type Communication (MTC), which is the cellular version of Machine-to-Machine (M2M) communication [1]. With a large number of IoT devices and the huge amount of data they generate, there must be a significant shift in both cellular and non-cellular wireless technologies to ensure the necessary level of connectivity. Hence, cellular networks, which have become exploited mostly by data traffic, are expected to achieve tremendous advances to keep up with these rising challenges. In particular, the coexistence of Human-Type Communication (HTC) and MTC has become inevitable [2]. In this regard, the Fifth Generation (5G) standard classifies the rising challenges to cellular communications into three substantial directions, namely, Ultra-Reliable and Low-Latency Communications (URLLC), Enhanced Mobile Broad-Band (eMBB), and massive Machine-Type Communication (mMTC) [3].

Ultra-Dense Network (UDN) is intuitively one of the most promising approaches to tackle such diverse requirements of massive connectivity and high throughput of the coexisting HTC and MTC. In a UDN, a massive number of low-power and low-cost Small Cells (SCs) are deployed with a density higher than that of the HTC users (HTCUs) [4]. With each HTCU associated with the closest SC, the cellular network acts as a mobile network following the HTCUs to remain as close as possible to them. With shorter serving distances between the SCs and the users/devices, we obtain enhanced performance in terms of achievable capacity. Besides, with the SC serving a smaller coverage area, more devices can be supported. Hence, within the UDN architecture, we aim at exploiting diverse available resources and technologies in 5G and Beyond (5GB) networks to tackle the above-mentioned challenges of higher capacity for the HTCUs and massive connectivity of the MTC devices (MTCDs).

We propose multiple-association of an HTCU to multiple SCs in order to achieve higher capacity by avoiding the bottleneck limitations of the backhaul capacities. By activating more SCs, more MTCDs can be supported. In addition, we investigate the possible gains in supporting computation-intensive applications by integrating edge computing servers within the SCs. Physical Layer Security protocols are also applied to secure the gigantic amount of data flying over the air. Applying Non-Orthogonal Multiple Access (NOMA) to support more MTCDs is considered as well. Finally, we utilize the Millimeter wave (mmWave) band to support MTC while considering wireless backhauling for the SCs. In general, we use tools from stochastic geometry to investigate and show the possible gains of deploying UDN along with the above-mentioned integrated technologies, propose other additional techniques and approaches to further enhance the achievable performance, and conclude by discussing the open problems for future research directions.

1.2 Motivation

UDN is a strong candidate for 5GB networks to satisfy the diverse requirements of future cellular networks. However, many challenges have to be considered and investigated to achieve full utilization of network resources. For example, providing a wired backhauling system such as optical fibers for the massive number of SCs may be practically infeasible and forces severe constraints on the achievable performance of UDNs [5]. Besides, the immense amounts of Inter-Cell Interference (ICI) found in UDNs may prevent the targeted gains of higher capacities from being achieved [4]. The massive number of MTCDs projected to coexist with HTCUs adds another dimension of complexity. With the recent increase in the number of data breaches due to the higher computational powers of illegitimate receivers, securing data transfer through only traditional cryptographic tools becomes questionable [6]. Finally, the increasing density of computation-intensive applications requires bringing the processing nodes to the edge of the network [7]. In addition to all of the aforementioned challenges, the large scale of UDNs in terms of the massive number of SCs and users/devices necessitates finding an alternative to the time-consuming simulations.

Hence, in this thesis, we aim at tackling the above-mentioned challenges accompanying the deployment of UDNs and supportive technologies. First, we use tools from stochastic geometry to formulate novel mathematical frameworks that provide closed or semi-closed analytical expressions of the average achievable performance against different system parameters. By doing so, we provide an alternative to the heavy time-consuming Monte-Carlo simulations and help network designers to better understand the network operation. We consider multiple-association of HTCUs to many SCs to benefit from the existing short distances between any HTCU and the set of the closest SCs. In doing so, we overcome the backhaul limitations, allow for parallel processing of computation-intensive applications through multiple Edge Computing Servers (ECSs), and exploit spatial diversity of the different links to secure transmitted data against illegitimate eavesdroppers.

In addition, to support a higher density of MTCDs, we apply NOMA to support multiple MTCDs on the same time-frequency resources. Also, we utilize the mmWave band to exploit the excess bandwidth available and the distinguished propagation characteristics of mmWave signals to mitigate the severe ICI found in UDNs.

1.3 Contributions

In this thesis and motivated by the above, we tackle multiple challenges of UDN deployment and explore the possible opportunities of integrating other supportive tools and/or technologies. The contributions of this thesis are summarized as follows.

Limited Backhaul Capacity and Coexisting HTC and MTC: We tackle the backhaul capacity limitations problem using multiple associations of SCs to an HTCU. By doing so, the required downlink high data rates of HTCUs can be split and transmitted through multiple SCs to match the available backhaul capacities. Besides, more SCs will be activated which allows for a higher number of MTCDs to be supported. It is important to note that the additionally activated SCs will serve the tagged HTCU on orthogonal bands, *i.e.,* no additional ICI will be introduced. In doing so, we study the effect of the limited backhaul capacity on the system while considering a fixed bandwidth. Also, we investigate the effect of multiple-association on MTC performance.

In particular, we show that an optimum *MultiCell* size exists which depends on the backhaul link capacities, density of SCs, and density of users. Besides, to efficiently share the available bandwidth among the multiple cells serving the same HTCU, we consider the different scenarios of how the HTCUs may be associated with SCs including those scenarios where a conflict may exist among the different users. In addition, we study the uplink performance of both HTC and MTC. For the MTC, by employing the proposed multiple-association scheme, we show the improvement in terms of both the density of supported MTCDs and achievable ASE. However, for uplink HTC, a single association to the closest

SC is considered sufficient since the required rates are much lower than the downlink scenario. To accurately investigate the performance of a UDN environment, we consider the SEPL model which takes into consideration the high probability of Line-of-Sight (LoS) communications. Using tools from stochastic geometry, we provide analytical expressions for the achievable ASE in HTC and MTC and the density of supported MTCDs.

Computation-Intensive Applications of HTCUs: We tackle the challenging growth of computation-intensive applications of HTCUs such as face recognition, virtual reality, and mobile augmented reality. In doing so, we exploit UDN and Multi-Access Edge Computing (MEC) to reduce the end-to-end processing delay experienced by the HTCUs. To the best of our knowledge, this is the first attempt to jointly consider edge computing with UDN along with multiple associations of SCs to the HTCUs in order to fully exploit the network resources.

In particular, each HTCU associates with multiple SCs equipped with ECSs and partially offloads its task to be processed in parallel at the ECSs and the local processor. We carefully mitigate the severe ICI experienced in UDNs by considering how adjacent HTCUs compete to associate with the same SC. Links that produce larger ICI and/or force significant offloading transmission delays are disconnected. Based on the relative densities between HTCUs and SCs, we use tools from stochastic geometry to develop an offline adaptive task division approach that further reduces the average end-to-end processing delay per user. Besides, we derive novel analytical expressions for the average end-to-end processing delay per user. The developed mathematical framework provides significant insights into the system and can be used to tune several parameters to achieve optimal performance.

Security Enhancement of HTC: To tackle data breaches, we adopt Physical Layer Security (PLS) and propose a secure multiple-association scheme in which the data traffic of an HTCU is split into different paths through the M closest cells. In doing, we exploit the spatial diversity among the different paths to increase the achievable secrecy

rate in the network. We show that exploiting the proposed multiple-association in a UDN environment can mitigate the effect of several deteriorating factors such as the increasing densities of both HTCUs and eavesdroppers.

NOMA and Excessive Density of MTCDs: We propose a system of hybrid NOMA/OMA scheme to support a massive number of MTCDs in a UDN environment. We divide the MTCDs into two groups according to their normalized Signal-to-Interference Ratio (SIR) compared to a predetermined SIR threshold value. Then, two users, one from each group, form a two-user NOMA cluster which can be extended to higher orders. We show that an optimum SIR threshold value can be chosen to maximize the obtained performance gain from the NOMA/OMA scheme compared to an only-OMA scheme. Using tools from stochastic geometry, we derive analytical expressions for the Area Spectral Efficiency (ASE) gain and show that densifying the network by more cells compromises the gain obtained from NOMA.

Millimeter Wave and Wireless Backhaul for Excessive Density of MTCDs: To fully exploit the available cellular network resources in 5GB, we jointly consider UDN and mmWave to support mMTC. By doing so, mmWave can help mitigate the severe ICI of UDNs thanks to the high penetration losses experienced in this frequency band. Besides, the gigantic amount of available bandwidths in the mmWave band helps to allocate a higher number of Resource Blocks (RBs) to the SCs that can be utilized to support more MTCDs per SC. Using tools from stochastic geometry, we provide analytical expressions for the achievable performance in terms of density of supported MTCDs, achievable SINR per MTCD, and capacity per SC.

Considering a practical use case, we study the performance under a limited wireless backhaul scenario while taking into account the correlated effects of different system parameters on both the access link (AL) and the backhaul link (BH). In doing so, we consider a massive Multiple-Input Multiple-Output (mMIMO) microwave backhauling system between the SCs and the Central Processing Units (CPUs) of the core network.

Besides, We propose a sub-optimal calculation of SC density, at which the achievable gains from the deployed SCs in terms of the limited capacity are maximized. Accordingly, we offer network operators a reasonable estimation of the density of SCs that should be activated. Moreover, we formulate a novel mathematical framework and derive analytical expressions for the density of supported MTCDs, the achievable SINR per MTCD, and the average capacity per SC (for both AL and BH) using tools from stochastic geometry and we derive a tight lower bound of the achievable limited capacity.

The above contributions have been published in [8, 8–10]. Other research works conducted during the tenure of this Ph.D. have been published in [11–13].

1.4 Thesis Organization

The rest of the thesis is organized as follows. In Chapter 2, we provide a concise background of the different topics and technologies included in the thesis. In Chapter 3, we consider multiple-association to overcome the challenge of backhaul capacity limitations in UDNs and to support higher densities of MTCDs. Chapter 4 tackles the challenge of computation-intensive applications of HTCUs through Edge Computing and multiple-association. Enhancing the HTC secrecy performance through multiple-association in UDNs is investigated in Chapter 5. In Chapter 6, we consider NOMA as a complementary approach for supporting more MTCDs. The gains associated with the utilization of the mmWave band while considering the wireless backhaul scenario are covered in Chapter 7. Finally, in Chapter 8, we draw the conclusions of the work conducted throughout this thesis and highlight open problems for possible future research directions.

Chapter 2

Background

2.1 HTC and MTC in 5G and Beyond Networks

The unprecedented growth in the number of connected users and/or devices and their traffic requires revolutionary technologies in cellular communications in order to cope with these requirements. According to [2], the number of connected devices by 2023 will be three times the global population. The fastest-growing mobile category between 2018 and 2023 will be Machine-to-Machine (M2M) communications. It will grow at a 19% Compound Annual Growth Rate (CAGR) or nearly 2.4 folds. This reflects an increase from around 6.1 billion devices in 2018 to approximately 14.7 billion devices by 2023. Within the M2M category, connected car applications will be the fastest growing category with 30% CAGR. In the same interval, smartphones will grow at a 7% CAGR (or 1.4 fold) reflecting the second fastest-growing category with an increase from 4.9 billion devices to 6.7 billion. Hence, it becomes clear that the coexistence of Human-Type Communication (HTC) and Machine-Type Communication (MTC), the cellular backbone of M2M communications, is inevitable [1, 12, 14].

Accordingly, tremendous advances in enabling technologies are needed in the next generations of cellular communications, fifth generation and beyond (5GB). In general,

5G classifies the rising challenges to cellular communications into three substantial directions, namely, Ultra-Reliable and Low-Latency Communications (URLLC), Enhanced Mobile Broad-Band (eMBB), and massive Machine-Type Communication (mMTC) [1,3]. For HTC, International Mobile Telecommunications-2020 (IMT-2020 Standard) requires enhanced peak data rates of 20 Gbits/sec [15]. For MTC, it is required to support a density of at least one million devices per km^2 [16]. Besides, computation-intensive applications such as virtual reality, face recognition, and augmented reality need to be addressed by bringing the processing servers to the edge of the network. Such architecture is referred to as Edge Computing [7] or Multi-Access Edge Computing (MEC) according to the European Telecommunications Standards Institute (ETSI) [17]. Also, with the frequent serious data breaches experienced recently, securing data has become more of a business risk rather than an information technology (IT) issue [2]. Hence, in this thesis, we aim at tackling some of the diverse requirements of 5GB use cases such as supporting coexisting HTC and MTC, handling computation-intensive applications of HTCUs, supporting a massive number of MTCDs simultaneously, and enhancing communication secrecy.

2.2 Ultra-Dense Network

Ultra-Dense Network (UDN) is considered a paramount candidate for providing the utmost growth of network capacity in 5GB [11]. In particular, a massive number of Small Cells (SCs) are to be deployed with low cost, low transmission power, and small coverage area [4]. Adding more cells, hence, reducing the serving area of each cell has always been the trait to improve the performance of cellular networks, especially in urban areas [4]. By doing so, the SCs get closer to the served users and/or devices. However, when the distances between the users and their serving/neighboring cells are shortened, the high probability of Line-of-Sight (LoS) should be taken into consideration. Besides, the significant sensed amounts of Inter-Cell Interference (ICI), which form severe challenges

to UDNs [12], must be tackled. Idle mode capability, where SCs with no associated users/devices are switched-off, can be seen as one solution to tackle the severe ICI problem [12, 18, 19]. It is noteworthy that densifying BSs was thought to have no effect on the signal-to-interference-plus-noise ratio (SINR). However, it was proven recently that this behavior will not continue beyond certain values of densities of SCs [20]. It is to be also noted that deploying SCs can be done by customers themselves within their premises. This can provide an optimal scenario in crowded areas to guarantee a certain quality of service (QoS).

In fact, many surveys and tutorials already exist covering the different aspects of UDNs. For example, [4] presents an extensive survey on UDNs, their modeling techniques such as stochastic geometry and game theory; performance metrics like coverage/outage probability, rate coverage/outage, average spectral efficiency, area spectral efficiency (ASE), network throughput, energy efficiency, and fairness among served users. In addition, the authors therein discussed different user association techniques like range expansion, dual connectivity, and multiple-association as well as the idle mode capabilities. The pioneering work reported in [21] discussed the behavior of ASE in a UDN environment with stretched exponential path loss (SEPL) model. This model takes into consideration the effects of short distances and blockages in order to provide accurate yet tractable analysis in contrast to the multi-slope path loss model. In the same scope, the authors in [22] provided asymptotic analysis for the ASE in UDN for a general small fading model and different feasible path loss models. Hence, in this thesis, we aim at exploiting the UDN architecture to tackle the raised challenges in 5GB towards supporting HTC and MTC.

2.2.1 Multiple-Association

In a UDN environment, the cellular network can be seen as a mobile network following the users. In other words, the serving SC will be always close to the user,

which significantly enhances the quality of the radio link. However, with the very high density of SCs, it becomes challenging to support all SCs by fiber links leading to limited backhaul capacities [5]. Hence, to fully exploit the potentials of UDNs, a Human-Type Communication User (HTCU) can be allowed to associate with multiple SCs [23–26]. It is to be noted that the density of the HTCUs is typically lower than that of the SCs [24]. By doing so, HTCUs can achieve higher data rates by exploiting the offered diversity of the different link qualities and the short distances between transmitters and receivers in a UDN. Also, the limitations forced on the backhaul link capacities can be mitigated [5,23]. It is noteworthy that dual connectivity is already considered by the third generation partnership project (3GPP) to boost user rates by receiving different streams of data from different sources, to achieve robustness against errors by receiving the same data from different sources, to enhance mobility support, or to provide seamless migration from 4G to 5G [27].

In addition to the above, the integration between UDN and MEC can benefit from multiple-association to tackle computation-intensive applications that can be partitioned during the run-time. Such applications are referred to as elastic applications [28]. In particular, partitioning is used to split the operational logic of the application into smaller sub-tasks [28]. Then, the sub-tasks are processed in parallel at the local processor of the HTCU and/or the multiple edge computing servers (ECSs) integrated within the multiple associated SCs. In a different scope, multiple-association can also be exploited to secure data traffic of HTCUs. In particular, the data traffic of the user can be split into different paths through the multiple associated SCs. Then, the spatial diversity among the different paths can be exploited to enhance the secrecy performance against existing passive eavesdroppers.

2.2.2 Wireless Backhaul

An inevitable bottleneck in the UDN performance is the backhauling system which is responsible for communicating the data between the SCs and the core network [23, 24]. Although providing backhaul links through optical fibers can provide a backhaul capacity of more than ten Gbps with very limited latency in the order of hundreds of microseconds [29], the accompanying cost and complexity would turn it into an infeasible solution [30]. Hence, wireless backhauling would be a promising alternative where SCs act as relays from the core network to the targeted users/devices and vice-versa, however, with its own accompanying challenges such as limited capacity. Besides, when adopting in-band self-backhauling, self-interference must be carefully managed between the Access Link (AL) and the Backhaul link (BH) [31, 32].

2.2.3 Stochastic Geometry

Stochastic geometry is one of the most commonly adopted tools to model UDNs [22, 33, 34]. In particular, the very high densities of SCs in UDN yield very complex and time-consuming Monte-Carlo simulations. Hence, providing closed or semi-closed mathematical expressions which capture the achievable performance in UDN against several system parameters is significantly needed. In other words, stochastic geometry offers the ability to study the average performance of large-scale networks with an infinite number of nodes while considering the effect of the ICI, a main limiting factor in the performance of UDNs [4].

From a mathematical perspective, the random deployment of the SCs by operators and customers as well as the locations of users/devices can be modeled by Point Processes (PPs). A PP is a set of random nodes existing in a one, two, or three-dimensional space. When the number of nodes follows a Poisson distribution, the PP is referred to as a Poisson PP (PPP) which is one of the most commonly adopted PPs for its mathematical tractability. Moreover, when the nodes are distributed independently and uniformly, the

PPP is referred to as a Homogeneous PPP (HPPP). Otherwise, When the locations of the deployed nodes are correlated, it becomes a Non-Homogeneous PPP. In stochastic geometry, a typical node is a node located anywhere (typically at the origin for simplicity) and reflects the average performance of all other nodes of the same type by applying Slivnyak's theorem [35]. In addition to the PPP, different applications may require different types of PPs. For example, in a Poisson cluster process (PCP), there are parent points that are distributed according to an HPPP while the *off-springs* generated for each parent point are distributed uniformly around each parent point and independently from the parent distribution. In this case, the set of all *off-springs* represents the PCP [36]. In this thesis, we use stochastic geometry to formulate mathematical frameworks that characterize the performance of the considered system models against several system parameters.

2.3 Edge Computing

With the explosive growth of computation-intensive applications moving to the cloud, tremendous costs would exist to transfer the massive amounts of data between the end nodes and the central processing clouds [7]. Recently, applications such as virtual reality, face recognition, and augmented reality are gaining ubiquitous popularity. The vast evolution of powerful mobile devices and communication infrastructure help to intensify the growth of such applications, which in turn, require further advancements in the supporting technologies [37]. As a consequence, the traditional cellular architectures may no longer constitute a feasible solution, and the need for MEC becomes inevitable [17]. By doing so, the end nodes (subscribers) would benefit from lower latency, and the backhaul links would become less congested. Besides, additional benefits can be acquired in terms of security compared to moving large amounts of data to a central cloud [7]. Hence, in this thesis, we make use of the MEC and integrate it within the UDN architecture to achieve enhanced performance for the HTCUs when processing computations-intensive

applications.

2.4 Millimeter Wave

Millimeter wave (mmWave) band can be seen as a promising candidate to tackle many challenges in 5GB networks [38, 39]. In particular, the tremendous amount of bandwidth available in the mmWave band can be exploited to tackle the unprecedented required data rates by complementing or even replacing some microwave bands like the sub-6 GHz band. Besides, more RBs can be allocated for each SC to support more connections/devices. Moreover, the short wavelengths of the mmWave signals allow a massive number of antennas to be deployed in small areas [40]. Having a massive number of antennas enhances the directional beamforming gain and helps to boost the received signal power and mitigate ICI. It is noteworthy that 5G Release-15 and Release-16 standardized by the 3GPP considered the mmWave band operation [41].

However, other factors have to be reconsidered, ahead of which is the blockage effect on signal propagation. The mmWave signals can travel for only a few kilometers or even less due to their short wavelengths and the accompanying severe path loss. In particular, the propagation characteristics of mmWave signals are distinguished by significant penetration losses in the received signal powers. However, with large densities of SCs in UDNs, there is a higher probability of shorter serving distances and Line of Sight (LoS) links between each device and its serving SC. On the contrary, as the distance increases, especially in a dense environment of blockages such as indoors, the probability of Non-LoS (NLoS) links significantly increases. In this thesis, we utilize the mmWave band in UDNs to mitigate the large amounts of sensed ICI and to support more MTCDs.

2.5 Non-Orthogonal Multiple-Access

Orthogonal multiple access (OMA) schemes have been efficiently deployed to share resources orthogonally in time, domain, or codes. In a different scope, multiple-input-multiple-output (MIMO) has been used to obtain spatial diversity. Then, Non-Orthogonal Multiple-Access (NOMA) has proven to add an extra dimension for efficient utilization of spectrum through power and/or codes. In power-domain NOMA (PD-NOMA), different power levels are used to differentiate the signals of different users. Then, successive interference cancellation (SIC) is used to decode the signal with higher power and then subtract it from the total signal, then, decode the lower power signal. For recent advances in research regarding NOMA, the readers may refer to [42] and the references therein. The authors clearly distinguish between dominant types of NOMA; power-domain NOMA and code-domain NOMA. In [43], a unified model for NOMA-enabled heterogeneous-UDNs was provided regarding user association and resource allocation.

NOMA has been considered extensively in academia and it was considered as a study item of 5G New Radio (NR) [44]. In NOMA [42], several users (two or more) share the same time-frequency-spatial resources simultaneously exploiting either the power domain or code domain. However, the earned gains were not significant enough to continue with NOMA as a work item for the NR and it was decided to be postponed for possible next generations in order to tackle the existing challenges [45]. Amongst the challenges obstructing the efficient deployment of NOMA is the NOMA clustering process [46]. In particular, which users will share the same resource, what is the optimal cluster size, and prior to these, what is the objective of a NOMA cluster, *i.e.,* maximizing the network throughput or achieving fairness among the different users? While SIC is responsible for mitigating the intra-cell or Inter-NOMA Interference (INI), a non-negligible complexity is added to the receiver. Also, imperfect SIC significantly degrades the gain obtained from adopting NOMA [47]. In the same scope, ordering the users in the SIC process based on either the instantaneous received power or on the mean received power (distance)

should be considered. For the former, *i.e.,* the instantaneous received power, Channel State Information (CSI) should be acquired which increases the overhead in the system. However, inaccurate or non-optimal performance may result from using distance-based ordering. Hence, in this thesis, we aim at exploiting NOMA to support more MTCDs without incurring additional costs on the network.

2.6 Physical Layer Security

Securing the gigantic amount of wireless information flying over the air has evolved from just an information technology (IT) issue to more of a business threat [2]. With the huge advances in computational capabilities and the increased scalability of networks, traditional cryptographic protocols may not be enough to secure wireless communications [6]. Hence, Physical Layer Security (PLS) protocols based on information-theoretic approaches are seen as a promising alternative to complement the traditional cryptographic ones [34]. In doing so, the inevitable distinctions among the channels of legitimate and illegitimate receivers are exploited to secure wireless communications [48]. With the introduction of UDNs, the distances between transmitters and legitimate receivers are shortened and the large-scale fading is significantly reduced. Such an environment existing in UDNs allows for optimal application of the PLS protocols that require better links for the legitimate receivers compared to the illegitimate ones [49]. In this thesis, we utilize PLS protocols within a UDN environment to provide an additional layer of security for HTC.

Chapter 3

Coexisting HTC/MTC in Limited-Backhaul Capacity UDNs with Multiple-Association

3.1 Introduction

The coexistence of Human-Type Communications (HTCs) and Machine-Type Communications (MTCs) is inevitable in future cellular communications [1,12,14]. However, a Machine-Type Communication Device (MTCD) should be handled differently compared to a Human-Type Communication User (HTCU) [11, 18]. In MTC, devices will communicate with each other with minimal human intervention. MTCDs with their small packet sizes, their massive numbers, and their required massive number of simultaneous connections impose significant challenges on the next generations of cellular networks.

While UDN can significantly enhance the radio link by shortening the distances between transmitters and receivers, backhaul links may impose practical capacity limitations. In particular, with the very high density of SCs, it becomes challenging to support them by fiber links leading to limited backhaul capacities [5]. Hence, multiple associa-

tions of SCs [24] can mitigate such limitations in the backhaul capacity, specifically, in the downlink. However, the tremendous downlink Inter-Cell Interference (ICI) found in a UDN must be mitigated by adopting idle mode capabilities of the SCs [12]. In other words, only an HTCU should be allowed to activate one or more SCs. Although exploiting the UDN environment requires MTCDs to be able to activate the closest SC as well, this may lead to cases where almost all SCs are activated due to the very high density of MTCDs.

In this chapter, we tackle two challenges of future cellular networks, namely, the backhaul capacity limitations in UDNs and the coexistence of HTC and MTC. Using multiple associations of SCs to an HTCU, the required downlink high data rates by HTCUs can be split among multiple SCs to match the available backhaul capacities. Besides, more SCs will be activated and a higher number of MTCDs can be supported. It is important to note that the additionally activated SCs will serve the tagged HTCU on orthogonal bands, *i.e.,* no ICI. Unlike the work in [24], we study the effect of the limited backhaul capacity on the system while considering a fixed bandwidth and we investigate the effect of the multiple-association scheme on the MTC performance. The main contributions in this chapter can be summarized as follows:

- We propose a multiple-association scheme for HTCUs under fixed bandwidth allocation and limited backhaul capacity to improve the achievable ASE. We show that an optimum *MultiCell* size exists which depends on the backhaul link capacities, density of SCs, and density of users.

- To efficiently share the available bandwidth among the multiple cells serving the same HTCU, we consider the different scenarios of how the HTCUs may be associated to SCs including those scenarios where a conflict may exist among the different users.

- We study the uplink performance of both HTC and MTC. For the MTC, by employ-

ing the proposed multiple-association scheme, we show the improvement in terms of both the density of supported MTCDs and achievable ASE. However, for uplink HTC, a single association to the closest SC is considered sufficient since the required rates are much lower than the downlink scenario.

- To accurately investigate the performance of a UDN environment, we consider the SEPL model which takes into consideration the high probability of LoS communications. Using tools from stochastic geometry, we provide analytical expressions for the achievable ASE in HTC and MTC and the density of supported MTCDs.

3.1.1 Related Works

Limited backhaul capacities

Backhaul capacity constraints (or fronthaul capacity constraints in a Cloud Radio Access Network (CRAN)) have been tackled in many works in the literature [30, 50–54]. For example, in [50], the authors investigated the effect of a limited fronthaul capacity on the downlink performance of a heterogeneous CRAN. In doing so, they considered a hybrid Millimeter-Wave (mmWave) and free space optical fronthaul links. It was shown that different fronthaul capacities associated to the Remote Radio Heads (RRHs) require different biasing factors of these RRHs to provide better coverage. Also, Integrated Access and Backhaul (IAB) has been considered in [55–57]. In [55], the authors considered mmWave bands and investigated both throughput and communication latency in an IAB scenario. In [56, 57], the authors analyzed the downlink rate coverage probability using mmWave for IAB under different bandwidth partitioning strategies. Alternatively, the works in [58, 59] analyzed the performance of a finite fronthaul capacity cell-free massive Multiple-Input Multiple-Output (MIMO) system in the downlink and uplink, respectively.

Coexistence of HTC and MTC

Although HTC and MTC have been tackled in many previous works, few works investigated their coexistence. For example, the authors in [60] investigated the performance of a single type of user (HTCU) for both uplink and downlink in MIMO systems. They compared between disjoint and user-centric clustering of cooperating base stations (BSs) to boost the achievable performance. However, the authors therein assumed unbounded backhaul capacity and a much higher density of users than BSs which are different from our system model regarding the HTCUs. In [61], the author considered different scenarios of user point processes served in the uplink. The author differentiated among three cases of high, moderate, and low user densities and showed how the relative densities between users and BSs may require different approaches in their mathematical analyses. Also, the authors in [62] investigated the asymptotic downlink performance of very high BS and user densities.

Considering the coexistence of HTC and MTC, the authors in [12] investigated the performance of downlink NOMA Area Spectral Efficiency (ASE). They showed therein the trade-offs between the two association schemes of SCs to MTCDs, namely, Connect-to-Active (C2A) and Connect-to-closest (C2C). The downlink coverage and average cell load have been studied in [18] in a similar scenario. Alternatively, the uplink NOMA performance of coexistent HTC and MTC has been investigated in [19]. In [63], the authors proposed a deep learning algorithm to optimize resource allocations between coexistent HTCUs and MTCDs within a cognitive radio framework. HTCUs were considered as the primary users while MTCDs were dealt with as the secondary users where NOMA was adopted to differentiate between their messages at the BS.

The rest of the chapter is organized as follows. In Section 3.2.1, we describe the system model and problem formulation. Section 3.3 provides the necessary analysis for both HTC and MTC and concludes by giving closed-form expressions for the different performance metrics. The obtained results and findings through Monte-Carlo simulation

and analytical expressions are reported and discussed in Section 3.4.

3.2 System Model

3.2.1 Spatial Distributions

We consider a UDN environment with a massive number of limited-backhaul capacity SCs serving HTCUs and MTCDs. In our system model, we consider a downlink scenario of HTC where each HTCU may associate to more than one cell. In parallel, we study the effect of multiple-association on the performance of uplink MTC. Besides, the performance of uplink HTC is investigated. Since the MTCs are usually directed in the uplink where sensors upload data to the network, the downlink MTC is omitted. Hence, the downlink MTC is omitted in this chapter. Such considered scenarios lie under the use cases of both Enhanced Mobile Broad-Band (eMBB) and massive Machine-Type Communication (mMTC) targeted in the 5G and beyond [11]. All of the SCs, HTCUs, and MTCDs are spatially distributed according to three independent Homogeneous Poisson Point Processes (HPPPs), Φ_s, Φ_h, and Φ_m with intensities λ_s, λ_h, and λ_m, respectively. In a UDN environment, the density of SCs is higher than the density of HTCUs, *i.e.,* $\lambda_s > \lambda_h$ [4]. However, this is not the case for the MTCDs where a heavily loaded regime is assumed such that $\lambda_m \gg \lambda_s$ coincides with the mMTC scenario even under the UDN assumption. For a practical scenario, only a fraction of the existing MTCDs will be active at a certain time instant [64]. Hence, we assume a fraction ρ_m of MTCDs will be active at a certain time instant which yields a thinned HPPP $\Phi_m^a \subset \Phi_m$ with density $\lambda_m^a = \rho_m \lambda_m$ for the active MTCDs. An illustration of the considered system model is shown in Fig. 3.1.

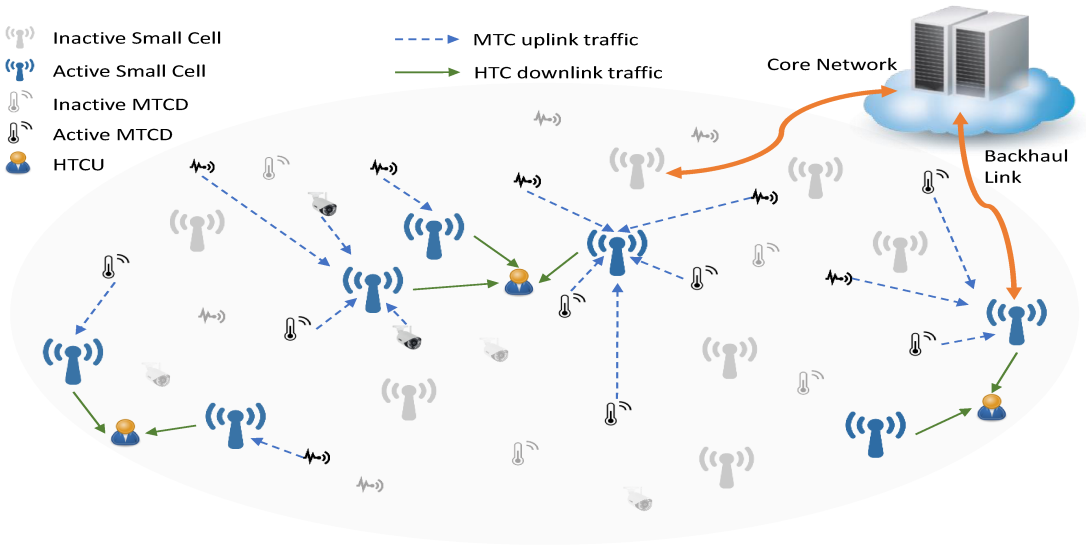


Figure 3.1: An illustration of the considered system model.

3.2.2 Channel Model and Interference

In the downlink, we consider ICI from only active cells where at least one HTCU is served by an active SC. Also, the achievable ASE comes solely from those active cells with their achievable rates being upper bounded by the backhaul link capacities. On the other hand, both uplink ASE and uplink ICI come solely from the supported HTCUs/MTCDs. We assume orthogonal frequency allocation for the HTCUs and MTCDs such that no mutual interference among the two types of users/devices. We also assume that each SC can support a maximum number of MTCDs equal to the number of available Resource Blocks (RBs) allocated for the MTC. If the number of associated active MTCDs to a specific cell is less than the available number of RBs allocated for MTC, then, those active MTCDs are randomly distributed over these RBs such that each MTCD exploits a single RB. Alternatively, if the number of associated active MTCDs is larger than the available number of RBs allocated for MTC, only a number of active MTCDs equal to the number of RBs can be supported by this cell.

All SCs, HTCUs, and MTCDs are equipped with single omnidirectional antennas. To reflect a UDN practical scenario, we adopt the SEPL model in which the signal

attenuates with distance d as $e^{-\alpha d^{\beta}}$ [21]. The SEPL model takes into account the high probability of LoS communications when the distances between transmitters and receivers are shortened. The two SEPL tunable parameters α and β are chosen in a way that captures different propagation scenarios [21]. For the multi-path fading, we assume Rayleigh fading channels where the channel gains are exponentially distributed with unit mean. Besides, we assume a block fading model such that the channel gain is fixed over a Transmit Time Interval (TTI) and changes independently from one TTI to another. In addition, we assume that all active transmitting nodes (SCs, HTCUs, and MTCDs) have infinitely backlogged packets to transmit.

3.2.3 Limited Backhaul Capacity

Taking into consideration the difficulties and challenges in supporting the massive number of SCs with sufficiently large capacities in the backhaul links, we assume limited backhaul capacities for the SCs in the downlink traffic [29]. However, for the uplink, both HTC and MTC traffic have relatively lower data rates. Hence, one does not need to consider specific limitations on the backhaul link capacities. To further illustrate, we assume that each SC is supported by a fixed limited normalized backhaul capacity ρ (bps/Hz) in the radio link, *i.e.,* the supported downlink rate by each SC per one Hz allocated in the radio link is upper bounded by ρ. This assumption reflects a more practical scenario where the provided backhaul capacity is proportional to the allocated bandwidth in the radio link [59]. Hence, the instantaneous achievable rate per cell per one Hz is

$$\hat{R} = \min(R, \rho) \tag{3.1}$$

where R is the instantaneous achievable rate per cell per one Hz in the radio link.

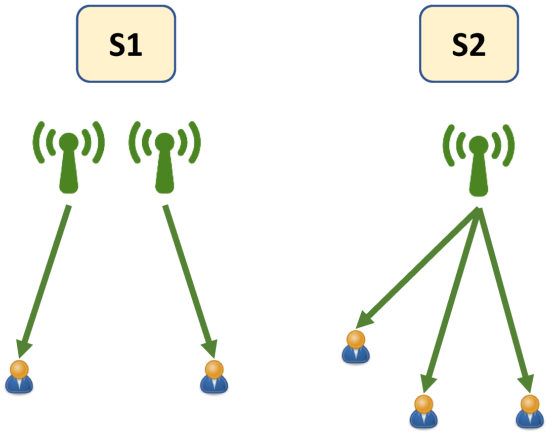


Figure 3.2: Different scenarios in single association scheme.

3.2.4 Downlink HTC and Multiple-Association

We refer to the conventional association scheme where each HTCU connects to and activates the closest SC as the single association scheme. In such a scheme, we expect two different scenarios illustrated in Fig. 3.2. In the first scenario, *S1*, an active SC serves exactly one HTCU while in the second scenario, *S2*, an active SC serves more than one HTCU. In the latter, we assume that the multiple users served by the same SC will share the available RBs orthogonally based on a Frequency Division Multiple Access (FDMA) approach. In doing so, the achievable data rate from a specific SC is divided among the multiple users such that no mutual interference exists among them. However, the considered SC will generate ICI to the neighboring cells over the whole bandwidth regardless of the number of served HTCUs.

For a double (multiple) association scheme, each HTCU connects to and activates the first and the second (up to M) closest SCs. We refer to the group of two (or more in case of multiple-association) cells as a *MultiCell*. Also, the set of the first closest SCs to the HTCUs are denoted by *tier-1*, the second closest set of SCs by *tier-2*, and so on. Low-ordered tiers refer to the near cells while higher-ordered tiers refer to further cells. Since the available spectrum is fixed and the frequency reuse factor is one, we assume that

the available frequency will be divided equally among the different tiers. It is noteworthy that the more bandwidth given to *tier-1*, the higher the spectrum efficiency will be. In other words, the highest spectrum efficiency can be achieved when the whole frequency band is allocated to *tier-1* (single association), however, the backhaul capacities will be the limiting factor.

In this chapter, we investigate the system performance under equal frequency allocation versus the *MultiCell* size. We assume that each HTCU is served via multiple traffic streams, where each stream corresponds to a tier (*i.e.,* an SC) on an orthogonal channel. Then, we calculate the sum-rate achievable by a certain HTCU which requires the HTCU to be equipped with M receivers with appropriately tuned filters to simultaneously decode traffic streams corresponding to the M tiers. In a practical scenario, some sort of calibration should be considered for the different achievable SINRs from the different tiers to use the appropriate corresponding encoding and decoding scheme. To overcome this problem, one can employ look-up tables or the Effective Exponential SNR Mapping (EESM) tool as in [65]. In our work, for mathematical tractability, we calculate the sum-rate of a user by summing the rates achievable from the different tiers which represents an upper bound of the actual achievable rates and area spectral efficiency.

3.2.5 Conflicting Small Cells

Similar to a single association, in a double association scheme, an active SC may serve exactly one user or more shown in Fig. 3.3. Different from a single association, an SC serving more than one user may be associated with users on different tiers (*S3*). We refer to those users which associate to the same cell but on different tiers as *conflicting users* and the cell as a conflicting SC. This will result in severe interference on the user(s) served from that cell on the higher-ordered (further) tier(s), *i.e.,* the interference will be higher than the desired signal. That is, each *tier-k* (set of small cells corresponding to the k^{th} closest SCs to the HTCUs) is allocated a sub-band of bandwidth $1/M$ where these

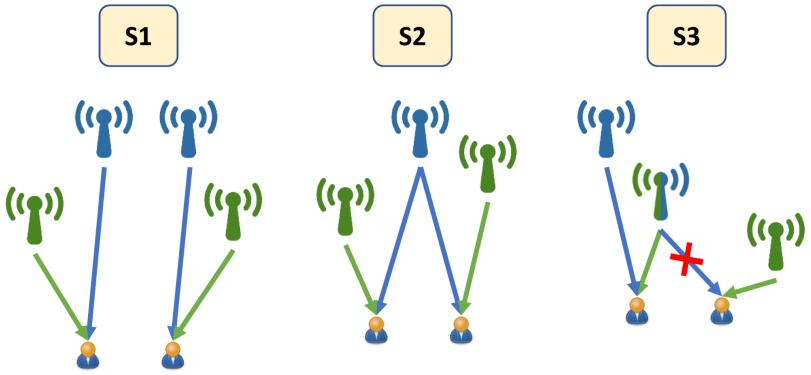


Figure 3.3: Different scenarios in double association scheme.

sub-bands are orthogonal and reused for all cells belonging to this tier. When a conflict occurs such that a specific cell is of *tier-k* for one user and *tier-m* for another and *tier-n* for a third user where $k < m < n$, then, the connections belonging to *tier-m* and *tier-n* for this specific cell are disconnected. In other words, the cell operates as *tier-k* and transmits only over the sub-band allocated to *tier-k*. Hence, other HTCUs that should have been served from this cell on *tier-m* and *tier-n* are no longer served on the corresponding frequencies. It is noteworthy that optimizing the frequency sub-band allocations to the different cells with conflict will require optimization over the whole network which may not be practically feasible and is out of the scope of this work. In our work, we switch off (disconnect) the higher-ordered tier(s) of those *conflicting users* as illustrated in *S3* in Fig. 3.3. Note that when an SC serves more than one user on the same tier, the achievable data rate is divided among multiple users similar to the single association scheme. It is important to mention that our proposed model still achieves higher ASE under the limited backhaul capacity constraints compared to a single association.

In multiple-association, a system with a *MultiCell* size M will have the active cells clustered around each HTCU. Hence, the active cells form a point process Φ_s^a thinned from the HPPP Φ_s. Fig. 3.4 illustrates how the active cells (dots) are distributed around the HTCUs for different sets of small cell densities. The Voronoi tessellation of the active cells is also illustrated where the coverage area of each active cell is represented by a

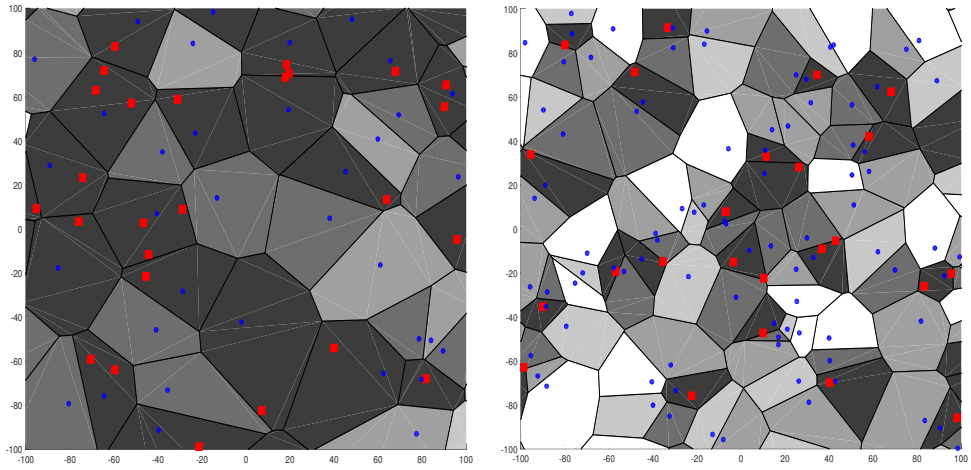
color reflecting the order of the tier on which the SC is activated. The darkest cells are those cells belonging to *tier-1* and brighter cells represent higher-ordered tiers. In case of a cell serving more than one HTCU on different tiers, the higher-ordered tier(s) will be disconnected as explained earlier. As a consequence, we can note from Fig. 3.4(a) that when the density of SCs λ_s is not much higher than that of the HTCUs λ_h, the majority of the cells are activated as low-ordered (dark-colored) tiers. As λ_s increases, more cells are activated on the different tiers while the cell coverage areas get smaller as clear from Fig. 3.4(b).

3.2.6 Uplink HTC

For the uplink HTC, a fraction ω of the available bandwidth is allocated for the HTC while the remaining $(1-\omega)$ fraction is allocated for the uplink MTC. Each HTCU transmits its data and this data is received and only decoded by the closest SC. As mentioned earlier, since the uplink HTC data rate is relatively lower than that of the downlink, the backhaul capacity limitations would not incur constraints on the achievable rates and thus are not considered. If two or more HTCUs are associated with the same SC in the uplink, they transmit their data to that cell sequentially based on a Time Division Multiple Access (TDMA) approach.

3.2.7 Machine-Type Communication (MTC)

Following the multiple-association scheme for the HTCUs, the active MTCDs will connect to the *closest-active* cell to upload their data. In doing so, the MTCDs will benefit from the additionally activated SCs to support their required massive number of simultaneous connections and to boost their achievable ASE. It is important to note that the additionally activated SCs in case of multiple-association do not generate severe ICI in the downlink since they operate on orthogonal frequency bands. For the uplink, however, a frequency reuse factor of one is adopted. In order to investigate the achievable gain, we



(a) $\lambda_s = 1,000$ cells/km^2 (b) $\lambda_s = 5,000$ cells/km^2

Figure 3.4: Different active cells realizations for different λ_s, where $\lambda_h = 500$ (HTCUs/km^2) and *MultiCell* size $M = 5$. Red squares indicate HTCUs and blue dots denote active cells. Dark and bright cells reflect low and high ordered tiers, respectively

focus on a practical scenario where each cell has a limited allocated bandwidth. In other words, we assume a fixed number of RBs, N_{RB}, per each active cell where each MTCD requires exactly one RB to transmit its data to the SC. As mentioned earlier, a fraction ω of these RBs is allocated to the HTCUs while the remaining $(1-\omega)$ fraction of RBs is allocated to the MTCDs. In addition, we assume that the MTCDs associated with a certain SC are randomly distributed over the available RBs in order to minimize the ICI for the uplink MTC. Hence, one may conclude that the number of supported MTCDs per cell is upper bounded by the number of allocated RBs for MTC.

The probability density function (pdf) of the cell size distribution of a randomly selected (typical Voronoi) cell in an HPPP was derived empirically using Monte-Carlo simulations in [66]. Furthermore, the authors in [67] exploited this cell size distribution to derive the probability mass function (pmf) of the number of users/devices per such typical cell as

$$f_N(n) = \frac{(\frac{3.5}{\lambda+3.5})^{3.5}}{\Gamma(3.5)} \left(\frac{\lambda}{\lambda+3.5}\right)^n \frac{\Gamma(n+3.5)}{\Gamma(n+1)}. \tag{3.2}$$

However, for the considered multiple-association scenario, the distribution of the cell sizes

will differ from one tier to another since the sizes of the neighboring cells are correlated. Hence, each tier should have its own cell size distribution as well as its corresponding distribution of the number of associated MTCDs. For mathematical tractability, we will assume that the distribution of the number of MTCDs in a randomly selected active cell follows (3.2) which is originally derived for an HPPP. We show later in the obtained results comparing simulations with the obtained analytical expressions that such approximation can be ignored when the relative densities λ_s/λ_h is not extremely large, *i.e.,* $\lambda_s/\lambda_h < 10$.

3.3 Performance Analysis

It is known that the main limiting factor for adopting a single association is the limited backhaul capacity. If the achievable rate in the radio link per cell per Hz is R and the backhaul capacity is ρ (bps/Hz), then, the actual achievable rate $\hat{R}$ will be as given in (3.1). Hence, we investigate the effect of the proposed multiple-association scheme under such backhaul capacity limitations by providing closed-form expression for the achievable downlink ASE for HTC. For completeness, we also study the uplink HTC performance. In addition, we proceed with the analysis to study the effects of the proposed multiple-association scheme on the MTC performance in terms of both the density of supported connections and the achievable ASE.

3.3.1 Multiple-Association

In the proposed multiple-association scheme, the number of active cells depends on the relative densities between SCs and HTCUs as well as the *MultiCell* size M. Since the HTCUs are uniformly distributed according to an HPPP, it is possible to have some HTCUs which are in close proximity to each other. Hence, these close-by HTCUs may be associated with common cells either on the same or on different tiers. Following our assumption illustrated in Fig. 3.3, any SC serving more than one user on different tiers

will be disconnected from the higher-ordered tier(s). As a result, we expect that the density of active cells on each tier will decrease as the tier order increases. It is to be noted that the density of active cells on a certain tier k, λ_k^a will be independent of the higher-ordered tiers. In other words, the density of active cells on *tier-k* is the same for any $M \geq k$. For example, the density of active cells on *tier-1*, λ_1^a, is the same whether we adopt single or multiple associations since in all cases they reflect the closest set of small cells to the HTCUs. Hence, using the pmf of the number of users per cell given in (3.2), the probability of cell activation on *tier-1* can be approximated as

$$p_1^a \approx 1 - f_N(0) = 1 - \left(\frac{3.5}{3.5 + \frac{\lambda_h}{\lambda_s}} \right)^{3.5}. \tag{3.3}$$

For higher order tiers, Lemma 3.1 gives the density of active cells in each tier.

Lemma 3.1 *The density of active cells on a specific tier-k,* $1 \leq k \leq M$ *can be approximated by*

$$\lambda_k^a \approx \mathcal{C} \left(\frac{3.5}{3.5 + \frac{k\lambda_h}{\lambda_s}} \right)^{3.5} \lambda_s, \tag{3.4}$$

$$\mathcal{C} = \left(\frac{3.5\lambda_s + k\lambda_h}{3.5\lambda_s + (k-1)\lambda_h} \right)^{3.5} - 1. \tag{3.5}$$

Proof: Let λ_k^{all} represent the density of all active cells on all tiers up to *tier-k*. The HTCUs are spatially distributed according to an HPPP such that each HTCU associates to the closest k cells ($k \leq M$). Alternatively, we follow the assumption in [26] and replace each HTCU with k users which are uniformly distributed such that each user associates only with the closest cell. The density of HTCUs becomes $k\lambda_h$ and the activation proba-

bility of all active cells up to *tier-k* is

$$\lambda_k^{all} \;=\; p_k^{all}\;\lambda_s \;=\; \left[1-\left(\frac{3.5}{3.5+\frac{k\lambda_h}{\lambda_s}}\right)^{3.5}\right]\lambda_s. \tag{3.6}$$

Next, we find the density of active cells on *tier-k* by subtracting $\lambda_k^{all}-\lambda_{k-1}^{all}$ which gives (3.4). ■

As the tier order increases in (3.6), the activation probability of all cells up to *tier-k* increases. However, from (3.4) and (3.5), the density of active cells on *tier-k* gets smaller.

3.3.2 Downlink Human-Type Communication

To investigate the performance of HTC, we assume a typical HTCU located at the origin which connects to and activates the closest M cells. This typical user reflects the performance of all users existing in the network under the same conditions according to Slivnyak's theorem [35]. Hence, the achievable rate by this typical user from the k^{th} closest SC while assuming it is being solely served by that SC reflects the average achievable rate per cell per one Hz on *tier-k*. From the HPPP distribution of the SCs, the probability that a circle of radius r centered at the typical user (origin) includes exactly k SCs is

$$\mathbb{P}(n=k;r)=\frac{\left(\pi\lambda_s r^2\right)^k}{k!}e^{-\pi\lambda_s r^2}. \tag{3.7}$$

Let r_k be the distance between the typical user and the k^{th} closest SC. Hence, the pdf of the distance between the typical user and the k^{th} closest SC is given by [35]

$$f_{r_k}(r)=\frac{2(\pi\lambda_s)^k}{(k-1)!}r^{2k-1}e^{-\pi\lambda_s r^2}. \tag{3.8}$$

The rate achievable by the typical user in the radio link from the k^{th} tier is given by

$$R_k = \log_2(1 + \gamma_k), \tag{3.9}$$

$$\gamma_k = \frac{P h_k e^{-\alpha r_k^{\beta}}}{I_k + \sigma^2} = \frac{h_k e^{-\alpha r_k^{\beta}}}{I'_k + \frac{\sigma^2}{P}}, \tag{3.10}$$

where γ_k is the instantaneous Signal-to-Interference plus Noise Ratio (SINR). P is the transmission power of the SC and h_k and r_k are the channel gain and the distance between the typical user and its serving cell on the k^{th} tier, respectively. α and β are the SEPL parameters and σ^2 is the power of the Additive White Gaussian Noise (AWGN). Furthermore, the term I'_k represents the normalized ICI at the typical user received from tier k. Since the bandwidth is shared orthogonally among the different tiers, the normalized ICI comes from only active cells on the same tier and is given by

$$I'_k = \sum_{j \in \Phi_k^a \setminus b_{k_0}} h_j e^{-\alpha r_j^{\beta}}, \tag{3.11}$$

where Φ_k^a is the set of all active cells on *tier-k*, b_{k_0} is the serving cell on *tier-k*, and h_j and r_j are the channel gain and distance between the typical user and the interfering cells belonging to *tier-k*, respectively.

Theorem 3.1 *In a multiple-association scheme with MultiCell size M, the average achievable rate per cell on tier-k with ideal (infinite) backhaul capacity, is given by*

$$\bar{R_k^M} = \begin{cases} \int_0^{\infty}\int_0^{\infty} \mathcal{A}(t,r)\ \mathrm{d}r\ \mathrm{d}t & \textit{; connected on tier-k} \\ 0 & \textit{; disconn. on tier-k} \end{cases}, \tag{3.12}$$

$$\mathcal{A}(t,r) = e^{-(2^{Mt}-1)\frac{\sigma^2}{P} e^{\alpha r^{\beta}}} \mathcal{L}_{I'_k}\Big((2^{Mt}-1)e^{\alpha r^{\beta}}\Big) f_{r_k}(r), \tag{3.13}$$

$$\mathcal{L}_{I'_k}(s) = \exp\left(-2\pi\lambda_k^a \int_r^\infty \left(\frac{1}{1+\frac{1}{s}e^{\alpha v^\beta}}\right) v \, \mathrm{d}v\right), \tag{3.14}$$

where λ_k^a *is obtained from (3.4) and* $f_{r_k}(r)$ *is given in (3.8).*

Proof: Appendix B.1. ■

Corollary 3.1 *For the special case when* $\beta = \frac{2}{n+1}$ *where* n *is a non-negative integer, the Laplace transform of the ICI given in (3.14) becomes*

$$\mathcal{L}_{I'_k}(\zeta e^{\alpha r^{\frac{2}{n+1}}}) = \exp\left(\sum_{i=0}^{n} \binom{n}{i} \alpha^i \frac{(n+1)\pi\lambda_k^a}{\alpha^{(n+1)}} \int_0^\infty \frac{-u^{n-i}}{1+\frac{1}{\zeta}\, e^u} \mathrm{d}u \cdot r^{\frac{2i}{n+1}}\right). \tag{3.15}$$

Proof: Appendix B.1. ■

It is worth noting that in (3.15), the distance r is not embedded in the integral as in (3.14) which simplifies the subsequent integral calculations in (3.12).

3.3.3 Limited-Backhaul Capacity

In the previous subsection, we assumed infinite backhaul capacity between the small cells and the core network. However, to account for the major challenge of limited backhaul capacity in UDNs, we evaluate herein the instantaneous achievable rate under such a scenario given in (3.1).

Corollary 3.2 *In a multiple-association scheme with MultiCell size* M *and normalized backhaul capacity* ρ *per one* Hz *in the radio link, the average achievable rate per cell on tier-k is*

$$\hat{R_k^M} = \begin{cases} \int_0^\rho \int_0^\infty \mathcal{B}(t,r) \, \mathrm{d}r \, \mathrm{d}t & ; \textit{connected on tier-k} \\ 0 & ; \textit{disconn. on tier-k} \end{cases}, \tag{3.16}$$

$$\mathcal{B}(t,r) = e^{-(2^{Mt}-1)\frac{\sigma^2}{P}e^{\alpha r^\beta}}\,\mathcal{L}_{I'_k}\Big((2^{Mt}-1)e^{\alpha r^\beta}\Big) f_{r_k}(r), \tag{3.17}$$

where $f_{r_k}(r)$ *is given in (3.8), and* $\mathcal{L}_{I'_k}(.)$ *is given in (3.14) or (3.15) where the latter corresponds to the special case of* $\beta = 2/(n+1)$.

Proof: Given ρ, the average achievable rate per cell on *tier-k* is given by,

$$\begin{aligned}
\hat{R}_k^M &= \mathbb{E}\left[\min\left(\rho, R_k^M\right)\right] \\
&= \int_0^\infty \mathbb{P}\left[\min\left(\rho, \frac{1}{M}\log_2(1+\gamma_k)\right) > t\right] \mathrm{d}t \\
&= \int_0^\infty \mathbb{P}\left[\rho > t, \frac{1}{M}\log_2(1+\gamma_k) > t\right] \mathrm{d}t \\
&= \int_0^\infty \mathbb{P}\left[\frac{1}{M}\log_2(1+\gamma_k) > t,\ t < \rho\right] \mathrm{d}t \\
&= \int_0^\rho \mathbb{P}\left[\gamma_k > 2^{Mt}-1\right] \mathrm{d}t.
\end{aligned} \tag{3.18}$$

and the rest of the proof follows as in Appendix B.1. ■

Corollary 3.3 *In a multiple association scheme with Multi-Cell size* M *and normalized backhaul capacity* ρ*, the average achievable ASE of the HTCUs,* $\mathcal{T}_h$ *is given by*

$$\mathcal{T}_h^d = \sum_{k=1}^{M} \lambda_k^a \int_0^\rho \int_0^\infty e^{-(2^{Mt}-1)\frac{\sigma^2}{P}e^{\alpha r^\beta}}\,\mathcal{L}_{I'_k}\Big((2^{Mt}-1)e^{\alpha r^\beta}\Big) f_{r_k}(r)\ \mathrm{d}r\ \mathrm{d}t, \tag{3.19}$$

where λ_k^a *is the density of active cells on tier-k given in (3.4), and* $\hat{R}_k^M$ *is the average achievable rate per cell on tier-k under limited backhaul constraints given in (3.16).*

Proof: The proof follows directly from (3.9) and the definition of ASE which is the total network achievable rate per one Hz within a unit area of one km^2. ■

3.3.4 Uplink Human-Type Communication

For completeness, we briefly show the performance of uplink HTC. For the uplink, each HTCU transmits its data while only the closest SC decodes and forwards this data it to the core network. The limited backhaul capacity in such a scenario does not impose notable constraints on the achievable rates since the uplink transmissions of both HTCUs and MTCDs (discussed in the next subsection) are relatively smaller than those of the downlink HTC. Hence, the scenario considered here reflects a single association scenario previously investigated in [19] and other works (*e.g.,* see [68,69]). A significant difference exists in the interference calculations between downlink and uplink. Since each HTCU connects to the closest SC, in the downlink, any interfering SC will be located further than the serving SC. However, in the uplink, although a typical SC will receive a corresponding signal from an HTCU located within its Voronoi cell, an interfering HTCU located in a neighboring cell can be closer than the served HTCU. Hence, dealing with interference in the uplink should be treated differently. In this regard, as in [61], we approximate the interfering HTCUs by a Poisson Point Process with intensity

$$\lambda_{ppp}(r) \approx \lambda_h(1 - e^{-\frac{12}{5}\pi\lambda_h r^2}), \tag{3.20}$$

where r is the distance from the receiving SC. Since the HTCUs reflect a low user density scenario, the occupied cells are relatively larger than the typical cell [61] which is referred to as the Crofton cell effect. However, the link distances between an SC and an associated HTCU follow the standard Rayleigh distribution given in (3.8) with $k = 1$. It is noteworthy that for the MTC scenario, the high density of MTCDs yields a similarly shaped link distance distribution but with an empirical correction factor that will be illustrated in the next subsection.

Theorem 3.2 *The average achievable uplink ASE of HTCUs when a fraction ω of the*

bandwidth is allocated for the HTC is given by

$$\mathcal{T}_h^u = \omega\lambda_1^a \int_0^\infty \int_0^\infty e^{-(2^t-1)\frac{\sigma^2}{P_0}e^{\alpha r^\beta}} \mathcal{L}_{I_h^{up}}((2^t-1)e^{\alpha r^\beta}) f_{r_1}(r)\ \mathrm{d}r\mathrm{d}t, \tag{3.21}$$

$$\mathcal{L}_{I_h^{up}}(s) = \exp\left(-2\pi\lambda_1^a \int_0^\infty (1 - e^{-\frac{12}{5}\pi\lambda_1^a z^2})(\frac{z}{1+\frac{1}{s}e^{\alpha z^\beta}})\ \mathrm{d}z\right) \tag{3.22}$$

with λ_1^a *given in (3.4) and* $f_{r_1}(r)$ *given in (3.8) by setting* $k = 1$.

Proof: Appendix B.1. ■

3.3.5 Machine-Type Communication

The active MTCDs form a thinned HPPP Φ_m^a such that $\Phi_m^a \subseteq \Phi_m$, and $\lambda_m^a = \rho_m\lambda_m$. These active devices are uniformly distributed over the area under study where each active MTCD connects to the closest active cell. The active cells, however, form another thinned point process from Φ_s with density λ_M^{all} given in (3.6). In this regard, we assume that each active cell has a limited number of RBs and that each MTCD requires exactly one RB to transmit its data. Hence, the number of supported devices by each active cell is upper bounded by the number of available RBs allocated for the MTC N_{RB} such that

$$N_{m,s}^{cell} = \min(N_m^{cell}, N_{RB}), \tag{3.23}$$

where N_m^{cell} is the number of associated MTCDs to a specific cell and $N_{m,s}^{cell}$ is the corresponding number of supported devices by this cell. Note that the available bandwidth B for uplink communications is divided orthogonally between the uplink HTC and uplink MTC such that the N_{RB} RBs exploits $(1-\omega)B$ of the bandwidth.

Density of Supported MTCDs

The density of supported MTCDs represents a key performance metric for MTC. Although MTCDs tend to upload small packets of data, it is challenging to support their required massive number of connections. Hence, increasing the density of supported MTCDs constitutes a major challenge that can be tackled by adopting the proposed multiple-association scheme. In doing so, as we increase the *MultiCell* size M, the density of active cells λ_M^{all} given in (3.6) increases. Hence, the ability of the network to simultaneously support a higher number of MTCDs significantly improves. It is to be noted that the activated cells on different tiers vary in both of their densities and cell size distributions as illustrated earlier in Fig. 3.4. Accordingly, the number of associated MTCDs per cell will vary from one cell to another and the average number of MTCDs per cell will also differ depending on the tier order. However, for the considered densities of SCs and HTCUs where $\lambda_s/\lambda_h < 10$, we assume that the randomly selected cell will have the same distribution for the number of devices per cell given in (3.2) regardless of the tier order. In particular, for mathematical tractability, when investigating the performance of MTC, we assume that the active cells are uniformly distributed according to an HPPP Φ_s^a with density λ_M^{all} evaluated from (3.6) and that their Voronoi cell sizes are Independently and Identically Distributed (i.i.d.). Following this assumption, the pmf of the number of MTCDs per an active cell N_m^{cell} can be approximated by (3.2) with $\lambda = \frac{\lambda_m^a}{\lambda_M^{all}}$.

Lemma 3.2 *The average density of supported MTCDs in a multiple-association scheme with MultiCell size M, number of RBs for MTC N_{RB}, MTCDs activation probability ρ, and densities λ_s, λ_h, and λ_m for the SCs, HTCUs, and MTCDs, respectively, can be approximated by*

$$\lambda_m^s \approx \mathcal{J}\lambda_M^{all}, \tag{3.24}$$

$$\mathcal{J} = \left(N_{RB} - \sum_{i=0}^{N_{RB}-1} \sum_{n=0}^{i} f_{N_m^{cell}}(n) \right), \tag{3.25}$$

$$f_{N_m^{cell}}(n) = \frac{\left(\frac{3.5\lambda_M^{all}}{\rho\lambda_m + 3.5\lambda_M^{all}}\right)^{3.5}}{\Gamma(3.5)} \left(\frac{\rho\lambda_m}{\rho\lambda_m + 3.5\lambda_M^{all}}\right)^n \frac{\Gamma(n+3.5)}{\Gamma(n+1)}. \tag{3.26}$$

where λ_M^{all} is given by (3.6).

Proof: Multiplying the density of supported devices per cell $\mathbb{E}[N_{m,s}^{cell}]$ by the density of all active cells λ_M^{all} gives (3.24). From (3.23), the density of supported MTCDs per cell is calculated as follows,

$$\begin{aligned}
\mathcal{J} = \mathbb{E}[N_{m,s}^{cell}] &= \sum_{i=0}^{\infty} \left(1 - F_{N_{m,s}^{cell}}(i)\right) \\
&= \sum_{i=0}^{\infty} \mathbb{P}\left[N_{m,s}^{cell} > i\right] \\
&= \sum_{i=0}^{\infty} \mathbb{P}\left[\min\left(N_m^{cell}, N_{RB}\right) > i\right] \\
&= \sum_{i=0}^{\infty} \mathbb{P}\left[N_m^{cell} > i, i < N_{RB}\right] \\
&= \sum_{i=0}^{N_{RB}-1} \mathbb{P}\left[N_m^{cell} > i\right] \\
&= \sum_{i=0}^{N_{RB}-1} \left(1 - \sum_{n=0}^{i} f_{N_m^{cell}}(n)\right),
\end{aligned} \tag{3.27}$$

where $f_{N_m^{cell}}(n) = \mathbb{P}[N_m^{cell} = n]$ is obtained from (3.23) by substituting $\lambda = \frac{\rho\lambda_m}{\lambda_M^{all}}$ leading to (3.26) which completes the proof. ■

Area Spectral Efficiency for MTC

Using stochastic geometry with the aid of Slivnyak's theorem [35], and without loss of generality, we assume a typical MTCD located at the origin. This typical device may or may not be supported by the network with certain probabilities. Thus, the achievable rate by this device as well as the interference it causes to other devices will both be zero when this typical device is not supported. Hence, when studying the MTC performance, we only consider the supported MTCDs whose density is given in Lemma 3.2. In this

regard, we assume the typical MTCD located at the origin which belongs to the thinned HPPP $\Phi_m^s \subseteq \Phi_m^a$ with density λ_m^s given in (3.24). The achievable rate by this MTCD on a specific RB is given by

$$R_m = \frac{1-\omega}{N_{RB}} \log_2(1+\gamma_m), \tag{3.28}$$

$$\gamma_m = \frac{P_m^0 h_m e^{-\alpha r_m^\beta}}{I_m + \sigma^2} = \frac{h_m e^{-\alpha r_m^\beta}}{I'_m + \frac{\sigma^2}{P_m^0}}, \tag{3.29}$$

where γ_m is the SINR, P_m^0 is the transmission power of the typical MTCD which is assumed fixed, and h_m and r_m are the channel gain and distance between the typical MTCD and the tagged closest active cell, respectively. I'_m is the normalized ICI at the tagged SC from all other supported MTCDs from neighboring cells transmitting their data on the same frequency band of the tagged RB.

Since the typical MTCD is being served on a single RB, and all RBs are orthogonal in frequency, the ICI is generated from only those MTCDs allocated the same RB. In order to minimize the ICI, we assume that the supported MTCDs are randomly assigned to the available RBs. When evaluating the ICI, we consider another thinned HPPP $\Phi_{m,N}^s \subseteq \Phi_m^s$ which corresponds to the supported MTCDs on a specific RB such that $\lambda_{m,N}^s = \frac{\lambda_m^s}{N_{RB}}$. Thus,

$$I'_m = \sum_{j \in \Phi_{m,N}^s \setminus d_0} h_j e^{-\alpha r_j^\beta}, \tag{3.30}$$

where h_j, r_j are the channel gain and distance between the tagged cell of the typical MTCD and the interfering MTCDs belonging to the neighboring cells, respectively. Taking into consideration the Crofton cell effect [61, 70] which implies that the cell containing the typical MTCD tends to have a larger size and more devices sharing the resources, the obtained expression would not change since we consider only the MTCDs which are supported and allocated an RB. However, as considered in the previous subsection, we approximate the interfering MTCDs by a PPP with density $\lambda_M^{all}(1-e^{-\frac{12}{5}\pi\lambda_M^{all} r^2})$. Different from the uplink HTC scenario, the serving link distance distribution in the MTC scenario

of a high user density regime should include an empirical correction factor of 13/10 in the density of serving cells [61]. Hence,

$$f_{r_m}(r) = 2a\pi\lambda_M^{all} r \; e^{-a\pi\lambda_M^{all} r^2}. \tag{3.31}$$

with $a = 13/10$ in traditional single association scheme and λ_M^{all} is given in (3.6).

Theorem 3.3 *In a multiple-association scheme with MultiCell size M and a number of available RBs per cell for MTC N_{RB} utilizing a fraction $(1-\omega)$ of the available bandwidth, the average achievable rate per MTCD is given by*

$$\bar{R}_m = \begin{cases} \frac{(1-\omega)}{N_{RB}} \int_0^\infty \int_0^\infty \mathcal{E}(t,r) \; \mathrm{d}r \; \mathrm{d}t & ;\textit{MTCD supported} \\ 0 & ;\textit{MTCD not supp.} \end{cases}, \tag{3.32}$$

$$\mathcal{E}(t,r) = e^{-(2^t-1)\frac{\sigma^2}{P_m^0} e^{\alpha r^\beta}} \mathcal{L}_{I'_m}((2^t-1)e^{\alpha r^\beta}) f_{r_m}(r), \tag{3.33}$$

$$\mathcal{L}_{I'_m}(s) = \exp\left(-2\pi \frac{\mathcal{J}}{N_{RB}} \lambda_M^{all} \int_0^\infty (1 - e^{-\frac{12}{5}\pi\lambda_M^{all} z^2}) (\frac{z}{1+\frac{1}{s}e^{\alpha z^\beta}}) \; \mathrm{d}z \right) \tag{3.34}$$

where $\mathcal{J}$ is the average number of supported MTCDs per active cell given in (3.25) and λ_M^{all} is the density of all active cells given by (3.6).

Proof: Following the same steps in the proof of Theorem 3.2, then,

$$\mathbb{P}[\gamma_m > \zeta] = \int_0^\infty e^{-\zeta \frac{\sigma^2}{P_m^0} e^{\alpha r^\beta}} \mathcal{L}_{I'_m}(\zeta e^{\alpha r^\beta}) f_{r_m}(r) \; \mathrm{d}r, \tag{3.35}$$

where r_m is the distance between the typical MTCD and its tagged cell. Let r' be the minimum distance from the tagged cell to the closest interferer. Since the Voronoi cells do not have uniform shapes, r' can be larger or smaller than r_m. However, we adopt the same approximation followed in [61] which approximates the interfering nodes by a PPP with

variable density. For a typical MTCD, the distribution of r_m (the distance to the closest active cell), following the same assumption of the active cells being uniformly distributed is given in (3.31). Following the same steps in deriving (3.22) yields (3.34). Finally, the average achievable rate per MTCD can be calculated from $\bar{R}_m = \frac{1-\omega}{N_{RB}} \int_0^\infty \mathbb{P}[\gamma_k > 2^t - 1]\mathrm{d}t$ which gives (3.32) and completes the proof. ■

Corollary 3.4 *In a multiple association scheme with Multi-Cell size M, the average achievable ASE of the MTCDs, $\mathcal{T}_m^u$, is given by*

$$\mathcal{T}_m^u = \lambda_m^s \frac{(1-\omega)}{N_{RB}} \int_0^\infty \int_0^\infty e^{-(2^t-1)\frac{\sigma^2}{P_m^0} e^{\alpha r^\beta}} \mathcal{L}_{I'_m}((2^t-1)e^{\alpha r^\beta}) f_{r_m}(r)\ \mathrm{d}r\ \mathrm{d}t \tag{3.36}$$

where λ_m^s is the density of supported MTCDs given in (3.24), $(1-\omega)$ the fraction of bandwidth allocated for MTC, N_{RB} the number of RBs allocated for MTC, and $f_{r_m}(r)$ is the pdf of the distance distribution between MTCDs and their serving SCs given in (3.31).

Proof: The proof follows directly from (3.32) and the definition of uplink ASE which is the total network achievable rate per one Hz within a unit area of one km^2. ■

3.4 Simulation Results

We report the impacts of different system parameters on the performance of both HTC and MTC. In doing so, we simulate different scenarios by generating different realizations of the point processes and averaging the performance over these realizations. Unless otherwise stated, we consider the following parameters in Table 3.1.

3.4.1 Simulation Setup

We first generate the spatial realizations of the three different HPPPs Φ_s, Φ_h, and Φ_m with their corresponding densities. Then, each HTCU associates with and activates

Table 3.1: Simulation Parameters

Parameter	**Symbol**	**Value**
Simulation Area	A	1×1 km^2
Spatial Realizations	N_m	500
Time Realizations	N_t	10
MultiCell size	M	1 to 5
Small Cell density	λ_s	$5,000\ SC/km^2$
Norm. backhaul capacity	ρ	$1,\ 2,\ \infty$
HTCU density	λ_h	$500\ HTCUs/km^2$
Uplink HTC BW fraction	ω	0.5
MTCD density	λ_m	$1,000,000\ MTCDs/km^2$
MTC activation probability	ρ_m	0.1
Full Tx Power	P	23 dBm
Noise Power	σ^2	-174 dBm/Hz
Available RBs for MTC	N_{RB}	10
SEPL parameters	α and β	$\alpha = 0.94$ and $\beta = 1/2$

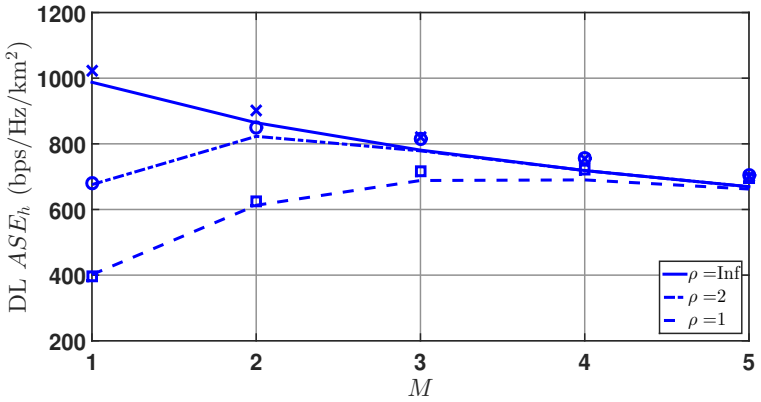
the closest M SCs. All the remaining SCs are switched to idle mode. If one cell is found to be serving more than one HTCU on different tiers, the higher tier is disconnected. For the uplink HTC, each HTCU uploads its data to the closest SC only utilizing a fraction ω of the available bandwidth. The remaining $1-\omega$ fraction available in the uplink is allocated for the MTC and divided into N_{RB} RBs. A fraction ρ_m of the MTCDs is randomly selected to become active and ready to transmit in correspondence to a real scenario where not all MTCDs are simultaneously active. Then, each of those active MTCDs associates with the closest active cell. If a cell has a number of associated MTCDs greater than the available RBs N_{RB}, a number of active MTCDs equal to N_{RB} are randomly selected to be supported. Alternatively, if the number of associated MTCDs is less than N_{RB}, then, they are randomly allocated RBs to transmit their data where each MTCD is allocated exactly one RB. When the achievable traffic of the downlink HTC on the radio link exceeds the available backhaul capacity ρ, the actual rate is then upper-bounded by this backhaul capacity. Finally, the ASE is calculated by summing all the achievable actual rates over the whole network in an area of one km^2 with a normalized bandwidth of one Hz.

3.4.2 Downlink Human-Type Communication

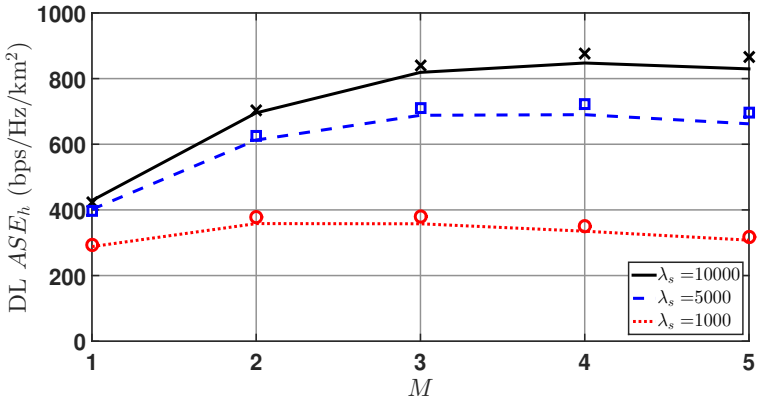
The achievable downlink ASE of HTCUs $\mathcal{T}_h^d$ versus the *MultiCell* size M is shown in Fig. 3.5 via both simulations and analysis (3.19). In Fig. 3.5(a), we show how the *MultiCell* size affects the downlink ASE of HTCUs under different normalized backhaul capacities. As evident from these results, with no limitation on backhaul capacity (i.e., $\rho = \infty$), a single association ($M = 1$) represents the best alternative for HTC. This is due to the high gain of the radio channel between the HTCU and its closest serving cell. However, for the practical case when the capacity of the backhaul link is limited, increasing M can benefit the HTCUs. One can notice from Fig.3.5 that different system parameters such as ρ and λ_s result in different values for the optimum *MultiCell* size $M = M^*$ at which the ASE is maximized. Note that for $M < M^*$, the backhaul capacity limitations yield lower ASE $\mathcal{T}_h^d$. For $M > M^*$, the channel degradation with a larger distance between the HTCU and the serving cell results in a lower ASE since the achievable rate on the radio links can be supported by the backhaul links of the M^* closest cells.

Fig. 3.6 shows the impact of increasing the backhaul capacity limit on the ASE for different *MultiCell* sizes. For the specified system parameters of λ_s and λ_h, we note that the double-association scheme ($M = 2$) achieves the highest ASE for the HTCUs when the normalized backhaul capacity ranges from around 1.5 to 3 bps/Hz. However, if $\rho > 3$ bps/Hz, the single association becomes the best alternative.

In Fig. 3.7, we show the impacts of varying the densities of both SCs and HTCUs on the performance of the HTC. One can notice that densifying the network enhances the achievable ASE for the HTCUs as shown in Fig. 3.7(a). However, when there exists backhaul capacity constraints, this achievable gain while adopting single association vanishes more rapidly as clear from the bottom curve. In Fig. 3.7(b), we show that the achievable ASE for the HTCUs also increases with the density of HTCUs λ_h under a fixed density of small cells $\lambda_s = 5000$ cells/km^2. However, when λ_h further increases under backhaul capacity limitations, one can notice that the ASE starts to decrease. Such behavior stems



(a) Different ρ and $\lambda_s = 5,000$ SCs/km^2.



(b) Different λ_s and $\rho = 1$ bps/Hz

Figure 3.5: Achievable ASE for HTCUs versus *MultiCell* size M for different small cell densities and normalized backhaul capacities with $\lambda_h = 500$ HTCUs/km^2. Lines indicate analysis while markers represent simulations.

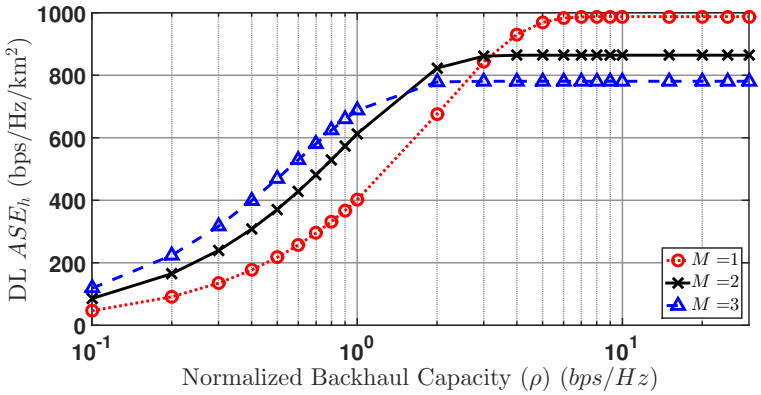
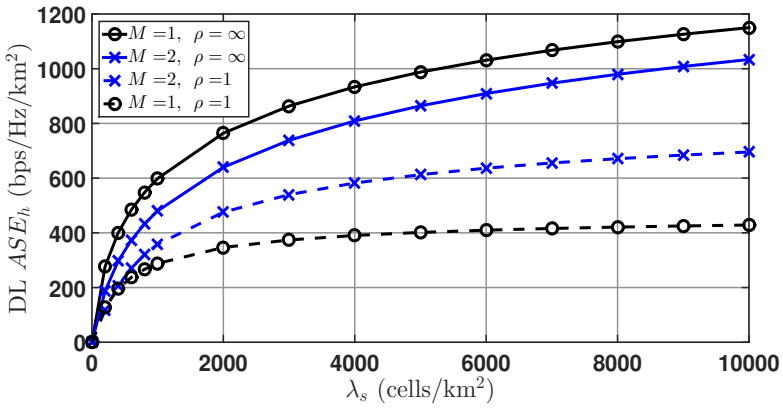
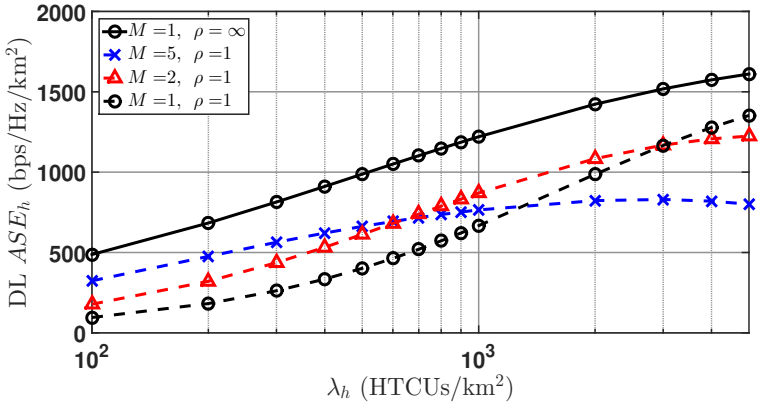


Figure 3.6: Achievable ASE for HTCUs versus normalized backhaul capacity ρ for different *MultiCell* sizes M, $\lambda_s = 5,000$ SCs/km^2 and $\lambda_h = 500$ HTCUs/km^2.



(a) $\lambda_h = 500$



(b) $\lambda_s = 5,000$

Figure 3.7: Achievable ASE for HTCUs versus small cell density λ_s and HTCUs density λ_h for different *MultiCell* sizes M and normalized backhaul capacities ρ.

from the more disconnected links of higher-ordered tiers arousing from the fact that more users will be sharing the same cells as previously discussed in Fig. 3.3. Doing so leaves the connected tiers with only a portion of the available bandwidth compared to the whole bandwidth in the case of a single association. However, when the density of HTCUs λ_h is relatively small compared to the density of small cells λ_s, multiple-association improves the achievable ASE under the same backhaul capacity limitations since more cells get activated.

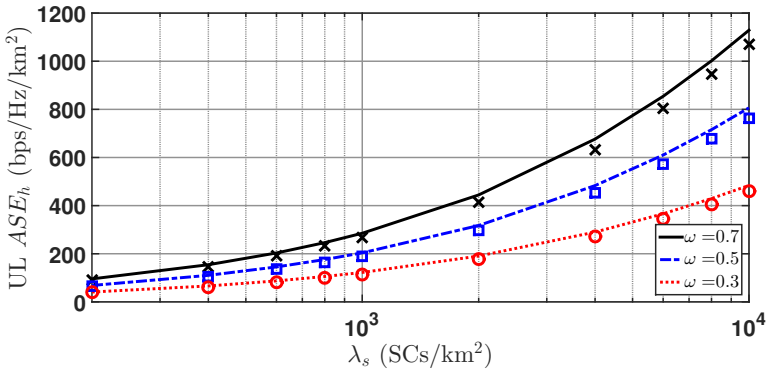


Figure 3.8: Achievable ASE for uplink HTC versus small cell density λ_s for different bandwidth allocation factor ω, and $\lambda_s/\lambda_h = 10$. Lines indicate analysis while markers represent simulations.

3.4.3 Uplink Human-Type Communication

Fig. 3.8 shows the achievable ASE for HTC in the uplink. The obtained results, comparing simulation results and the obtained analytical expression given in Theorem. 3.2, show that for a fixed ratio between λ_s and λ_h, increasing both densities significantly increases the achievable ASE. Besides, it is also clear that increasing the bandwidth fraction allocated for HTC ω yields higher uplink ASE for the HTCUs. It is noteworthy that multiple-association in the downlink does not impose any impacts on the achievable ASE in the uplink since HTCUs upload their data only to the closest SC.

3.4.4 Machine-Type Communication

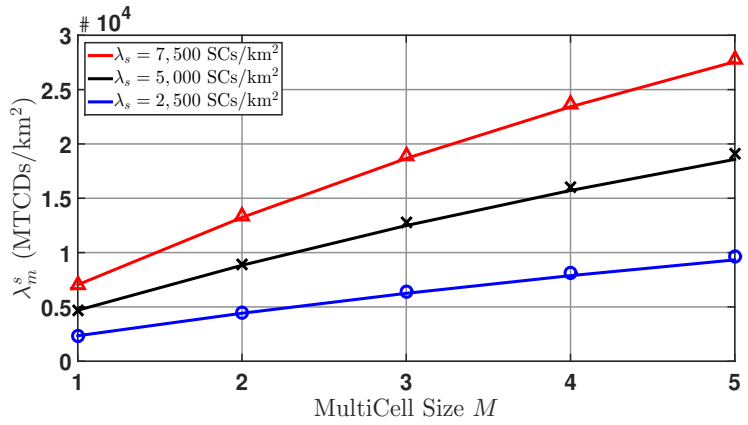
In this subsection, we show the gains that can be attained by the MTCDs from adopting the proposed multiple-association scheme for the HTCUs. In Fig. 3.9, we show through both simulations and analyses, the impact of increasing the *MultiCell* size M on both performance metrics of MTC, namely, the density of supported MTCDs λ_m^s and the achievable ASE $\mathcal{T}_m^u$ per 1 Hz allocated for MTC. As mentioned in the previous section, for the heavily loaded regime of MTC, the serving link distances are distributed according to (3.31) with $a = 13/10$ in case of single association [61]. For the $M > 1$,

the value of a should be smaller to compensate for the larger distances experienced on average for the MTCDs. We set $a = 13/(10 + \varepsilon)$ where ε increases with λ_s/λ_h and this increment depends on M as well. A good fitting formula for a is found empirically as $\varepsilon = 0.16\ (\log_2(b))^c/c;\ b = \lambda_s/2\lambda_h,\ c = \sqrt{\log_2 M}$. Clearly, one can notice from Fig. 3.9 that increasing M yields significant gains in both λ_m^s and $\mathcal{T}_m^u$ under different densities of small cells and HTCUs. Besides, one can notice from Fig. 3.10(a) that the higher the *MultiCell* size is, the higher the achievable gain is in the ASE of MTC with the increasing density of SCs λ_s.

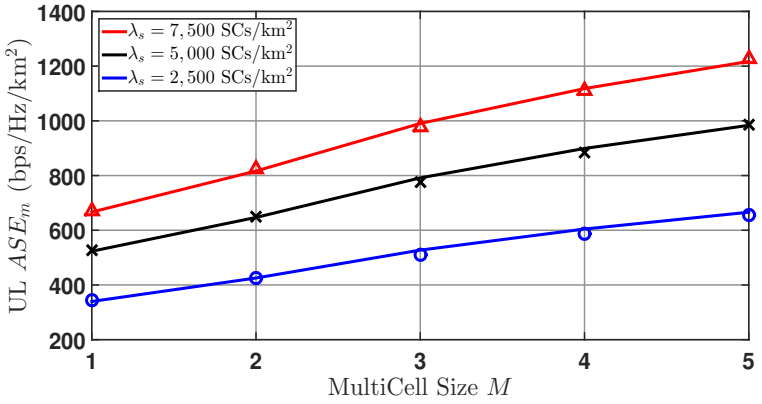
When the density of active MTCDs λ_m^a is small, the achievable ASE is also small as shown in Fig. 3.10(b). This occurs since the available RBs by all active cells can support those active MTCDs while plenty of resources are not being utilized. However, one can note that the achievable ASE is still higher with larger *MultiCell* sizes M. When M increases while the ratio λ_s/λ_h is fixed, more cells are activated which results in shorter distances between the active MTCDs and the active SCs to which they are associated. Since all supported MTCDs transmit with fixed power, shortening the distances between the MTCDs and their serving SCs yields a higher Signal-to-Interference Ratio (SIR) and achievable ASE. As λ_m^a increases, the ASE also increases until all the available RBs are utilized, then, the ASE saturates when no additional MTCDs can be supported. For higher M, there are more cells activated and the number of supported devices gets larger, hence, higher ASE is achieved.

3.5 Summary

We investigated the gains achieved from associating a Human-Type Communication User with multiple Small Cells in its vicinity. These gains are mainly obtained under limited backhaul capacity constraints that can be found in a UDN environment where fiber links require high deployment costs and time. Also, we show that the number



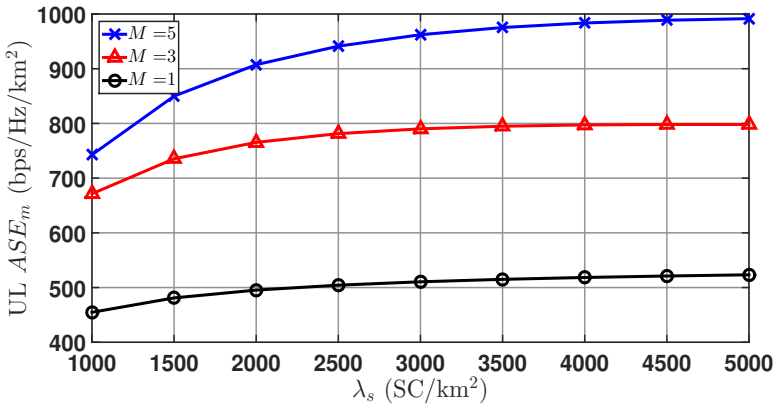
(a) Density of supported MTCDs (λ_m^s)



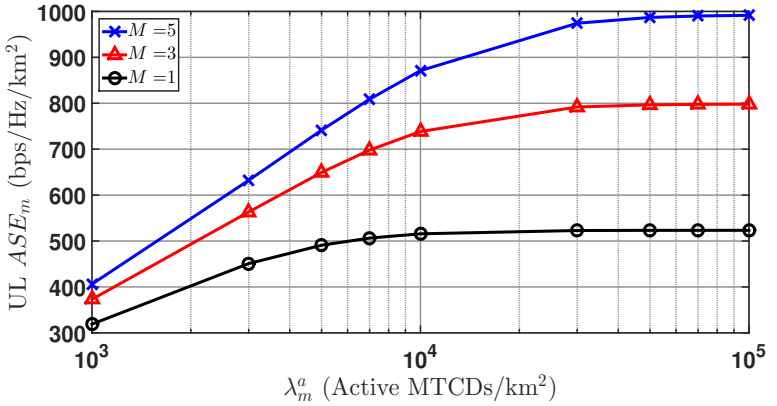
(b) ASE of MTCDs ($\mathcal{T}_m$)

Figure 3.9: Density of supported MTCDs and ASE of MTC versus M for different densities of SCs and HTCUs where $\lambda_s = 10\lambda_h$, and active MTCDs density $\lambda_m^a = 100,000$ (devices/km^2). Lines indicate analysis while markers represent simulations.

of associated cells can be optimized for different system parameters such as backhaul capacities and cells and users' densities. For practical scenarios of UDNs, the stretched exponential path loss model is adopted. In parallel, we investigated the gains achieved by the coexisting MTCDs. By adopting multiple-association, more cells are activated and a higher density of MTCDs can be supported. Following the proposed multiple-association scheme, we showed how UDN can be useful in supporting the different use cases targeted in 5G and beyond such as Enhanced Mobile Broad-Band and massive Machine-Type Communication. In this regard, we derived a mathematical framework analysis using tools from stochastic geometry to obtain analytical expressions for the achievable Area Spectral



(a) $\lambda_m^a = 100,000$ (devices/km^2)



(b) $\lambda_s = 5,000$ (cells/km^2)

Figure 3.10: ASE of MTC versus small cell density λ_s and active MTCDs density λ_m^a for different *MultiCell* sizes (M) with HTCUs density $\lambda_h = 500$ (users/km^2).

Efficiency under backhaul capacity limitations. Further extensions to this investigation could be resource allocation optimization in terms of frequency, power, and *MultiCell* size.

Chapter 4

Computation-Intensive Applications Supported by Edge Computing and Multiple-Association in UDNs

4.1 Introduction

With the explosive growth of computation-intensive applications moving to the cloud, tremendous costs would exist to transfer the massive amounts of data between the end nodes and the central processing clouds [7]. As a consequence, the traditional cellular architectures may no longer constitute a feasible solution, and the need to bring the processing servers to the edge of the network becomes inevitable. By doing so, the end nodes (subscribers) would benefit from lower latency, and the backhaul links would become less congested. Such architecture is referred to as Edge Computing [7] or Multi-Access Edge Computing (MEC) according to the European Telecommunications Standards Institute (ETSI) [17]. Besides, additional benefits can be acquired in terms of security compared to moving large amounts of data to a central cloud [7].

Hence, in this chapter, we tackle the challenging growing computation-intensive

applications of HTCUs in cellular networks such as face recognition, virtual reality, and mobile augmented reality. In particular, we exploit UDN and MEC to reduce the end-to-end processing delay experienced by the HTCUs. To the best of our knowledge, this is the first attempt to jointly consider edge computing with UDN along with multiple-association of SCs to the HTCUs in order to fully exploit the network resources. We use tools from stochastic geometry [71] to derive novel analytical expressions that can be used accurately and efficiently to evaluate the desired network performance against several system parameters without resorting to the time-consuming Monte-Carlo simulations. The main contributions can be summarized as follows:

- We jointly consider MEC and UDN with multiple-association to fully exploit the network's resources. In particular, each HTCU associates with multiple SCs equipped with ECSs and partially offloads its task to be processed in parallel at the ECSs and the local processor. Significant reduction in the average delay per user is achievable conditioned on the proper selection of the (*MultiCell* size).

- We carefully mitigate the severe ICI experienced in UDNs by considering how adjacent HTCUs compete to associate with the same SC. Links that produce larger ICI and/or force significant offloading transmission delays are disconnected.

- Based on the relative densities between HTCUs and SCs, we use tools from stochastic geometry to develop an offline adaptive task division approach that further reduces the average end-to-end processing delay per user.

- Using tools from stochastic geometry, we derive novel analytical expressions for the average end-to-end processing delay per user. The developed mathematical framework provides significant insights into the system and can be used to tune several parameters to achieve optimal performance.

4.1.1 Related Works

Interested readers may find plenty of works that jointly consider UDN with MEC to improve the quality of service offered to the served users. The main objective of MEC deployment varies between minimizing the energy consumption of the battery-limited users and/or minimizing their task processing delays. In this regard, existing works either develop optimization algorithms targeting the mentioned objective(s) or formulate a mathematical framework using tools from stochastic geometry to investigate the effect of different system parameters on achievable performance. It is noted that for feasibility concerns, those works developing optimization algorithms always consider a limited number of nodes, *i.e.,* SCs and users. In contrast, the main advantage of adopting stochastic geometry is the ability to study the average performance of large-scale networks with an infinite number of nodes while including the effect of the Inter-Cell Interference (ICI), a main limiting factor in the performance of UDNs [4].

Focusing on MEC and UDN with the support of multiple-association, task partitioning techniques play a critical role in achieving the required gains. Computation-intensive applications that can be partitioned during the run-time are referred to as elastic applications [28]. In particular, partitioning is used to split the operational logic of such applications into smaller sub-tasks [28]. In the field of parallel and distributed processing, one can find several examples of divisible computation-intensive applications [72]. It is noteworthy that a given application may be partitioned at different levels such as methods, threads, classes, or tasks [72]. However, the way the task is divided is beyond the scope of our work where we focus only on the total transmission and processing delays after the task has been partitioned. Interested readers may refer to [28] for detailed investigations on different partitioning techniques.

Examples of divisible applications include but are not limited to optical character recognition (OCR), face recognition, environment recognition, matrix computations, and signal processing applications. For example, in environment recognition, an HTCU

may use the device's camera to capture the surrounding environment. Then, the image/video frame can be partitioned into smaller sub-frames and be parallelly processed. The combinations of possible environments reported back for each sub-frame can be used to determine the exact environment. The same analogy can be applied to OCR where images of handwritten or printed text can be partitioned into distinct slices to be processed in parallel and transformed into digitized text [72]. Another example is object recognition where videos taken by cameras of autonomous vehicles are used to identify an object [73]. Several works in the literature considered arbitrarily divisible tasks for computation offloading and parallel processing, in particular, when the task is divided between the local processor and only one associated Edge Computing Servers ECS [74–77].

In [78], the authors proposed an algorithm to optimize task offloading decision and channel allocation in order to minimize energy consumption under a limited delay tolerance constraint. The simulated network consisted of one macro base station (MBS) and multiple SCs with one MEC server located at the MBS. The authors therein assumed an atomic task which cannot be partitioned, hence, the task is either computed locally or offloaded to the MEC server. The authors assumed optical fiber links between the SCs and the MBS and neglected the delays over these links. The work in [79] also considered atomic tasks and proposed a heuristic task offloading algorithm to optimize the objective function of weighted sum delay and energy consumption. In doing so, multiple MEC servers were considered and the allocation of computing resources, channels, and upload power were considered for optimization. Similarly, in [80], the authors investigated a preliminary model of offloading computation-intensive atomic tasks from the IoT devices to ECSs within a heterogeneous UDN environment. The authors proposed a heuristic greedy offloading scheme where an IoT device is allowed only one out of three computation schemes, namely, local processing, offloading to an ECS connected to the macro base station, or offloading to an ECS connected to an SC. The proposed algorithm aims at minimizing the energy consumption of the IoT devices while satisfying the processing

time constraints.

Unlike the above works, Jing *et al.* [81] investigated a scenario where smart devices split and offload their tasks simultaneously through multiple Radio Access Technologies (RATs) to multiple ECSs to be processed in parallel. In particular, the authors proposed an adaptive task splitting and resource allocation (communications and computation resources) algorithm to maximize the minimum averaged utility of the devices. Using extensive simulations, they showed that the proposed algorithm improves the fairness among the multiple devices in terms of the number of processed tasks. In the same scope, the authors in [82] considered a scenario where each user splits and offloads its task to multiple SCs equipped with ECSs. In this, the authors proposed a multi-agent deep deterministic policy gradient to jointly optimize task partitioning and uplink power control to maximize the long-term system utility which includes execution delay, and energy consumption. In both works [81, 82], a limited number of SCs were considered with respect to users which forces limitations on the available computational resources.

On the other hand, Ko *et al.* [83] investigated the performance of communication latency and computation latency using tools from stochastic geometry, queuing theory, and parallel computing. In particular, they provided scaling laws to quantify the trade-offs in terms of network latency, connectivity, and stability. In their considered system model, the authors assumed that each access point equipped with ECS has its serving zone with a fixed radius while each user associates with at most one ECS to offload its task. Those users outside of the serving zone of any ECS remain inactive. Similarly, the authors in [84] considered a scenario of a massive number of mobile users with computation-intensive tasks that can be either computed locally or offloaded to an access point with a co-located ECS whenever the local computation buffer is full. The average end-to-end latency is derived by integrating tools from stochastic geometry and queuing theory. The obtained results therein show how the local computing capabilities of mobile users can improve the performance of the network and give useful insights for network provisioning

and planning. It is noteworthy that the works in [83,84] did not consider task partitioning or parallel processing due to the higher density of users than the density of SCs.

Different from the above, Mei *et al.* [77] have recently considered a scenario of parallel computation where the user's task is split and parallelly processed between the local processor and the ECS. However, the authors therein considered only a single association scheme under two orthogonal frequency division multiple access (OFDMA) strategies. Using tools from stochastic network calculus and a Poisson Cluster Process (PCP) model, the authors modeled and analyzed the delay performance in a multi-cluster cellular network. In addition, the authors proposed a bisection search-based strategy for efficient data offloading.

The rest of the chapter is organized as follows. In Section 4.2, we describe the system model. Then, Section 4.3 provides the required mathematical analysis under two task partitioning approaches, namely, equal and adaptive task division. The Monte-Carlo simulation setup and the obtained results through both simulations and analytical expressions are reported and discussed in Section 4.4.

4.2 System Model

4.2.1 Network and Channel Model

We consider a cellular network where HTCUs exist in a UDN environment. The massively deployed SCs of the UDN are equipped with high computational capacity ECSs. Hence, the HTCUs attempt to offload their computation-intensive tasks, whenever possible, to these ECSs in order to reduce their end-to-end processing delays calculated from the instant when the task is generated until when its processing is completed. Thanks to the very high density of the SCs (and co-located ECSs) compared to the density of HTCUs [4], each HTCU is allowed to associate with multiple SCs (up to M) [23, 24] to communicate, split, and offload tasks to be processed in parallel. For a fair comparison,

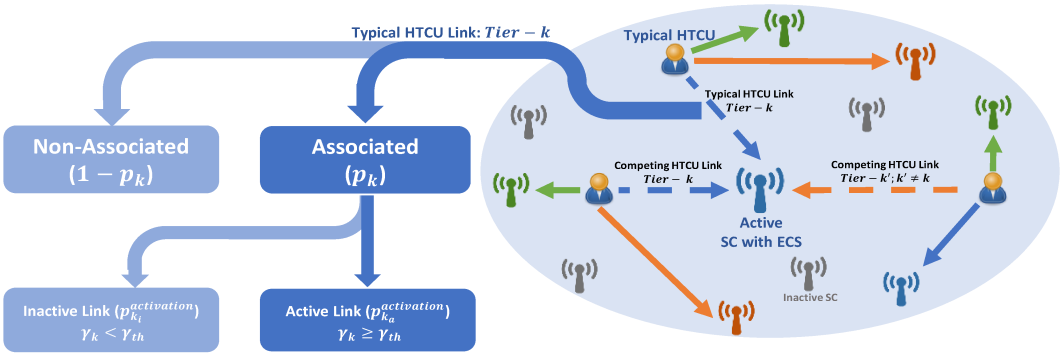


Figure 4.1: An illustration of the system model showing the association and activation concepts on a specific *Tier-k*

.

we divide the available bandwidth orthogonally and equally among the M tiers of SCs such that each tier is allocated W/M Hz where W is the total available bandwidth available for task offloading. By doing so, we guarantee zero interference among the different tiers. The motivation behind multiple-association is to exploit the available resources of UDNs for parallel, hence, achieving shorter delays. From a communication perspective, multiple-association can also provide higher data rates [24] and mitigate the limited backhaul capacities [23]. However, in this work, we focus on computation offloading and processing delay performance rather than communication performance.

We assume that all SCs (equipped with ECSs) and HTCUs are spatially distributed according to two independent Homogeneous Poisson Point Processes (HPPPs) Φ_s and Φ_h with densities λ_s and λ_h, respectively. Although repulsion between SCs is important in cellular network planning which raises concerns on the HPPP model of SCs, fortunately, we are more interested in calculating the aggregate interference rather than the spatial interactions between SCs [85]. In this regard, the offered analytical tractability by the Poisson point processes (PPP) model cannot be ignored given its simple probability generating functional (PGFL) expression [85]. Moreover, the HPPP version is stationary and spatially ergodic which further simplifies the analysis when applying Slivnyak's theorem where the average performance becomes independent from the observation location of the

typical node [85]. Such an assumption is widely adopted in the literature where the works in [21,24,64,68,86–88] considered UDN environment and assumed that the SCs and users are spatially distributed according to two independent HPPPs.

We assume that each HTCU $u \in \Phi_h$ can associate with up to M SCs to partially offload its computation-intensive task after proper partitioning [28]. We refer to the set of the closest SCs to the HTCUs as *Tier*-1, the set of the second closest SCs to the HTCUs as *Tier*-2, and so on up to *Tier*-M. Whenever a specific SC has more than one associated HTCU, only one user is selected by the SC to be served. In particular, when the different users are associated on different tiers, the link of with a higher-ordered tier (corresponding to an on-average further HTCU) is disconnected and denoted as non-associated. When two or more users are associated with the same SC on the same tier, only one user is randomly selected to be served and the remaining ones are non-associated. Hence, when we select the *MultiCell* size to be M, this means that each HTCU may associate with up to M SCs. However, some of these M SCs may not be associated due to competition from other HTCUs. Moreover, for an already-associated SC to an HTCU, the link between them is either active or inactive. Such link activation depends on whether the achievable signal-to-interference plus noise ratio (SINR) of this link satisfies the minimum required SINR-threshold γ_{th} or not. An illustration of the system model and the process of association and activation of certain *Tier*-k, $1 \leq k \leq M$ is shown in Fig. 4.1 where the probabilities of *Tier*-k association (p_k) and activation ($p_k^{activation}$) are defined in the next Section.

We adopt the traditional path loss model in which the signal power attenuates with the distance r according to $r^{-}\eta$. We assume that all channels suffer from Rayleigh fading in which the channel power gain is exponentially distributed with unit mean. We also assume that all nodes are equipped with single omnidirectional antennas.

4.2.2 Task Offloading Model

The task of an HTCU $u \in \Phi_h$ can be defined as (B_u, C_u) where B_u is the task transmit-size in bits and C_u is the required computational capacity in CPU cycles to process the task. For a tractable analysis, the task (B_u, C_u) is assumed identical for all users [83]. Also, each HTCU is equipped with a local processor with computational capacity f_{loc} (cycles/s). Hence, the processing delay experienced when the whole task is computed locally is

$$t_{loc} = \frac{C_u}{f_{loc}}. \tag{4.1}$$

Similarly, the ECSs integrated within the SCs each have computational capacities of f_{ecs} (cycles/s). If the whole task of the HTCU is processed at one of the ECSs, the processing delay becomes

$$t_{ecs} = \frac{C_u}{f_{ecs}}. \tag{4.2}$$

It is intuitive to assume that $f_{ecs} > f_{loc}$ which yields $t_{ecs} < t_{loc}$. However, when a task is offloaded for processing, the two-way transmission times of the task between the HTCU and the ECS should be included in the total delay.

We assume computation-intensive applications that can be arbitrarily partitioned during the run-time such as OCR [72], face recognition, or environment recognition. To exploit the full potentials of UDNs, we adopt multiple-association such that a computation task of certain HTCU can be arbitrarily partitioned to be processed in parallel at the local processor and all connected ECSs to this HTCU, known as partial offloading [89, 90]. By doing so, the experienced task processing delay per user can be found as the maximum of delays experienced over all connected links as well as the local processor. The delay over a connected (associated and active) link is composed of three parts; the offloading transmission time, the processing time at the ECS, and the transmission time of the processed results back to the HTCU. Typically, the processed results of a task have very small sizes compared to the task itself. Besides, the downlink rates are usually higher than

their counterparts in the uplink. Hence, the transmission time of the processed results back to the HTCU is neglected [84].

In general, we assume that both task transmit-size B_u and task required computational capacity C_u are partitioned proportionally, *i.e.,* with a partition $\alpha_k B_u$ of the task requires $\alpha_k C_u$ CPU cycles for processing. Such an assumption is acceptable for applications like OCR and environment recognition. Intuitively, we begin by assuming the computation-intensive task to be partitioned into equal parts. Afterward, we propose an adaptive task division that can significantly shorten the average experienced delay per user. The task processing delay experienced by an HTCU u over a *Tier-k*, $k \in \{1, 2, ..., M\}$ link can be expressed as

$$t_k^u = t_{k,tran}^u + t_{k,proc}^u = \alpha_k^u \left(\frac{B_u}{R_k^u} + t_{ecs} \right), \tag{4.3}$$

where $t_{k,tran}^u$ is the offloading transmission time, $t_{k,proc}^u$ is the processing time at the ECS, α_k^u is the fraction of the computational task allocated to *Tier-k*, and R_k^u is the uplink transmission rate achievable on *Tier-k*.

For equal task division, $\alpha_k^u = 1/\omega_u$ where $\omega_u = 1 + N_t$ is the number of the computational task divisions to be processed at the local processor and the connected ECSs such that N_t, $0 \leq N_t \leq M$ denotes the number of connected tiers (associated and active). The achievable instantaneous rate by user u on *Tier-k* is given by

$$R_k^u = \begin{cases} \frac{W}{M} \log_2(1 + \gamma_k^u) & ; \; Tier - k \text{ is connected} \\ & \quad \text{(associated and } \gamma_k^u \geq \gamma_{th}) \\ 0 & ; \text{ otherwise} \end{cases}, \tag{4.4}$$

where γ_k^u is the achievable instantaneous SINR by user u on *Tier-k*. For a given *Tier-k*, the achievable SINR is mainly governed by the received signal power at the tagged *Tier-k* SC of the HTCU, the sensed ICI at the tagged *Tier-k* SC of the HTCU on the frequency

band allocated to *Tier-k*, and the thermal noise at the receiver, hence, given as

$$\gamma_k^u = \frac{P_{u,k} h_u {r_{u,k}}^{-\eta}}{\sum_{j \in \Phi_{h,k} \setminus u} P_{j,k} h_j {r_{j,k}}^{-\eta} + \sigma^2}. \tag{4.5}$$

$P_{u,k}$ is the uplink transmission power of HTCU u on *Tier-k*, h_u and $r_{u,k}$ are the channel power gain and the distance between HTCU u and its tagged SC on *Tier-k* (the k^{th} closest SC), respectively, and η is the path loss exponent such that $\eta > 2$ to reflect an urban environment. On the other hand, $\Phi_{h,k}$ is a thinned HPPP from Φ_h that includes all HTCUs transmitting on *Tier-k* and causing interference to the tagged *Tier-k* SC of HTCU u. Hence, h_j and $r_{j,k}$ reflect the channel power gains and distances between all interfering HTCUs $j \in \Phi_{h,k} \setminus u$ and the tagged *Tier-k* SC of HTCU u, respectively. Finally, σ^2 is the noise power density.

Note that the uplink transmission powers of HTCUs on different tiers $P_{u,k}$ may change from one tier k to another or from one HTCU u to another. However, we assume $P_{u,k}$ to be fixed for all HTCUs and all tiers while being large enough to meet the required receiver's sensitivity, *i.e.,* $P_{u,k} = P_0 \ \forall \ u, \ k$. To justify this assumption, first, let the transmit power varies from one tier to another while it remains fixed for all users transmitting on the same tier, i.e. $P_{u,k} = P_k \ \forall \ u$. Then, the change (gain/loss) in the desired signal power would be neutralized by the change in the undesired interference power given that the considered network is interference-limited which is the case in a UDN environment [4]. The matching between simulations (noise considered) and analysis (noise neglected) in section 4.4 confirms that the considered scenario is interference-limited. Second, if the transmit power varies from one HTCU to another, then, a power allocation algorithm should be developed to optimize the achievable network performance towards a predetermined objective such as minimizing the average delay of all HTCUs or maximizing the fairness among the different HTCUs. However, such an approach lies beyond the scope of this work which aims at investigating the average performance using tools

from stochastic geometry to avoid extensive time-consuming simulations that would force additional undesired delay given the large scale of the UDN environment considered in this work. It is noteworthy that such optimization problems of either power allocation or channel allocation may be considered in future works while taking into consideration the additional incurred delay of the optimization process.

4.3 Performance Analysis

The main objective of this section is to derive an analytical expression for the average end-to-end processing delay per HTCU. Following the two considered task partitioning approaches, equal and adaptive task division, the average delay per user is derived for each approach separately. The different conditions of the links of different tiers specify the distinct possible combinations that an HTCU may experience. Hence, we start by calculating the association probability of each *Tier-k* as a function of the tier order, density of SCs, and density of HTCUs. Then, we calculate the link activation probability on all tiers both individually and jointly. It is noteworthy that the joint link activation probability is inevitable to proceed accurately with the analysis. To do so, the joint distance distribution among different tiers must be considered, which significantly complicates the analysis. Finally, we enumerate all possible combinations that an HTCU would experience, calculate the probability of each combination, find its corresponding delay, and average over all possible combinations.

4.3.1 Association Probability

First, to find the association probability on each *Tier-k*, $k \in \{1, 2, ..., M\}$, we assume a typical HTCU located at the origin with a typical tagged *Tier-k* SC. According to Slivnyak's theorem [35], when the SCs are distributed according to a PPP, the achievable performance of the typical HTCU reflects the performance of all HTCUs in the network.

In the adopted multiple-association scheme, the typical HTCU attempts to associate with and activate the closest M SCs. For a single-association scheme, the SC activation probability is acquired using an empirical formula derived in [67]. This formula is widely used in the literature as in [9, 11, 13, 91]. Moreover, under the considered multiple-association scheme and for a certain $k \leq M$, we assume that each user is replaced by k users that are uniformly and randomly distributed. However, with each user associating with and activating only one SC [23, 24]. By doing so, we compensate for the additionally activated SCs such that for a given k, we substitute the density of users by $k\lambda_h$ instead of λ_h when calculating the SC activation probability [26],

$$p_k^{all} \approx 1 - \left(\frac{3.5\lambda_s}{3.5\lambda_s + k\lambda_h} \right)^{3.5}, \tag{4.6}$$

where p_k^{all} is the SC activation probability on all tiers including *Tier*-1, $Tier-2, ...$ and up to *Tier-k*. The corresponding density of active SCs is $\lambda_k^{all} = p_k^{all}\lambda_s$. To obtain the SC activation probability on a specific *Tier-k*, $1 < k \leq M$, we subtract the SC activation probability of all preceding tiers up to $k-1$ from the SC activation probability of all current tiers up to k [23],

$$p_k^a = p_k^{all} - p_{k-1}^{all}. \tag{4.7}$$

Intuitively, for $k = 1$, $p_1^a = p_1^{all}$ where $p_0^{all} = 0$.

Lemma 4.1 *Association Probability: the probability that HTCU is associated with a Tier-k SC*

$$p_k = p_k^a \times \frac{\lambda_s}{\lambda_h} \tag{4.8}$$

where p_k^a is the density of active cells on Tier-k given in (4.7). Proof: Given that the density of HTCUs is λ_h, (4.8) directly follows. ■

4.3.2 Link Activation Probability

Individual and joint links' activations are determined by the achievable SINRs on the considered links with respect to a predetermined threshold γ_{th}. Hence, we begin by deriving the individual Complementary Cumulative Distribution Function (CCDF) of the achievable *Tier-k* SINR γ_k^u given in (4.5). For a simpler notation, we set $u = 0$ to denote the typical HTCU, and is omitted afterward. Also, note that the thinned HPPP $\Phi_{h,k}$ denoting the interfering HTCUs that are associated with *Tier-k* SCs has a corresponding density $p_k\lambda_h$ with p_k given in (4.8).

The distance distribution between the typical user and its closest SC can be obtained from the null probability of the HPPP and is given by $f_{r_1}(r_1) = 2\pi\lambda_s r_1 e^{-\pi\lambda_s r_1^2}$. However, in general, the distance distribution between the typical user and its k^{th} closest SC is given by [92]

$$f_{r_k}(r_k) = \frac{2(\pi\lambda_s)^k}{(k-1)!} r_k^{2k-1} e^{-\pi\lambda_s r_k^2}, \tag{4.9}$$

while the joint distance distribution of the closest k neighbours is given by [92]

$$f_{r_1,r_2,\ldots,r_k}(r_1, r_2, \ldots, r_k) = (2\pi\lambda_s)^k r_1 r_2 \ldots r_k e^{-\pi\lambda_s r_k^2}. \tag{4.10}$$

Lemma 4.2 *Individual Link Activation: in a multiple-association scheme with MultiCell size M and for $1 \leq k \leq M$, the individual CCDF of the achievable Tier-k signal-to-interference ratio (SIR) of the typical HTCU is*

$$\mathbb{P}\left(\gamma_k \geq \zeta_k\right) = \frac{1}{\left(1 + Q_k(\zeta_k)\right)^k}, \tag{4.11}$$

$$Q_k(\zeta_k) = p_k^a \zeta_k^{\frac{2}{\eta}} \int_{\zeta_k^{-2/\eta}}^{\infty} \frac{1}{1 + u^{\frac{\eta}{2}}} \mathrm{d}u, \tag{4.12}$$

where p_k^a is given in (4.7) and the link activation probability is obtained by setting $\zeta_k = \gamma_{th}$.

Proof: See Appendix B.2. ■

Using Lemma 4.2, the probability of *Tier-k* link activation and/or inactivation is

$$p_{k_a}^{activation} = \mathbb{P}(\gamma_k \geq \gamma_{th}), \tag{4.13}$$

$$p_{k_i}^{activation} = \mathbb{P}(\gamma_k < \gamma_{th}) = 1 - \mathbb{P}(\gamma_k \geq \gamma_{th}). \tag{4.14}$$

4.3.3 Joint Link Activation Probability and Conditional Joint CCDF

To accurately proceed with our mathematical analysis, the dependence among the different tiers should be considered. Hence, one needs to obtain the joint probability of link activation and/or inactivation on different tiers. In other words, the probability that the links on *Tier*-1, *Tier*-2, ..., and *Tier-k* are simultaneously following specific combination of activation and inactivation. When doing so, we must consider the joint distribution of the distances between the typical HTCU and the closest up to *Tier-k* SCs which is given in (4.10). For example, the joint link activation/inactivation probabilities on a two-tier scenario, *i.e.,* *Tier*-1 and *Tier*-2 are

$$p_{1_a,2_a}^{activation} = \mathbb{P}\left(\gamma_1 \geq \gamma_{th},\ \gamma_2 \geq \gamma_{th}\right), \tag{4.15}$$

$$\begin{aligned} p_{1_a,2_i}^{activation} =& \mathbb{P}\left(\gamma_1 \geq \gamma_{th},\ \gamma_2 < \gamma_{th}\right) \\ =& \mathbb{P}\left(\gamma_1 \geq \gamma_{th}\right) - \mathbb{P}\left(\gamma_1 \geq \gamma_{th},\ \gamma_2 \geq \gamma_{th}\right), \end{aligned} \tag{4.16}$$

$$\begin{aligned} p_{1_i,2_a}^{activation} =& \mathbb{P}\left(\gamma_1 < \gamma_{th},\ \gamma_2 \geq \gamma_{th}\right) \\ =& \mathbb{P}\left(\gamma_2 \geq \gamma_{th}\right) - \mathbb{P}\left(\gamma_1 \geq \gamma_{th},\ \gamma_2 \geq \gamma_{th}\right), \end{aligned} \tag{4.17}$$

$$p_{1_i,2_i}^{activation} = 1 - \mathbb{P}\left(\gamma_1 \geq \gamma_{th}\right) - \mathbb{P}\left(\gamma_2 \geq \gamma_{th}\right) + \mathbb{P}\left(\gamma_1 \geq \gamma_{th},\ \gamma_2 \geq \gamma_{th}\right), \tag{4.18}$$

where the subscript k_a denotes that link k is active while k_i denotes inactive. Hence, we need to find the joint CCDF of γ_1 and γ_2. In the next theorem, we give a general expression for the joint CCDF of the SIRs over multiple tiers.

Theorem 4.1 *Joint Link Activation: the joint CCDF of the achievable SIRs over multiple tiers* $..., i, j, k; 1 \leq ... < i < j < k \leq M$ *is given by*

$$\mathbb{P}\left(..., \gamma_i \geq \zeta_i, \gamma_j \geq \zeta_j, \gamma_k \geq \zeta_k\right) = \frac{1}{\left[1 + Q_k(\zeta_k)\right]^{k-j}\left[1 + Q_j(\zeta_j) + Q_k(\zeta_k)\right]^{j-i}...}, \tag{4.19}$$

$$\mathbb{P}\left(\gamma_i \geq \zeta_i, \gamma_j \geq \zeta_j, \gamma_k \geq \zeta_k\right) = \frac{1}{\left[1 + Q_k(\zeta_k)\right]^{k-j}} \times \frac{1}{\left[1 + Q_j(\zeta_j) + Q_k(\zeta_k)\right]^{j-i}} \times \frac{1}{\left[1 + Q_i(\zeta_i) + Q_j(\zeta_j) + Q_k(\zeta_k)\right]^{i}}. \tag{4.20}$$

where $Q_k(\zeta_k),\ \forall\ k \in \{1, 2, ..., M\}$ *is given in (4.12) and the joint link activation probability (all active) is obtained by setting* $\zeta_k = \gamma_{th}$. *Other joint activation/inactivation probabilities can be obtained using axioms of probability as in (4.16)-(4.18).*

Proof: See Appendix B.2. ■

In the following, we derive an expression for the conditional joint CCDF of SIRs which is necessary for calculating the average delay in the following subsections. In particular, given that certain tiers are connected (associated and active), we find the conditional joint CCDF of the achievable SIRs on these tiers while the states of all remaining tiers are also known in terms of association and activation. For example, $\mathbb{P}\big(\gamma_j \geq \zeta_j, \gamma_k \geq \zeta_k | ..., \gamma_j \geq \gamma_{th}, \gamma_k \geq \gamma_{th}, ...\big)$. To illustrate, let us begin with a single tier, *i.e.*, $\mathbb{P}\big(\gamma_k \geq \zeta_k | \gamma_k \geq \gamma_{th}\big)$. It

is straightforward that if $\zeta_k \leq \gamma_{th}$, then $\mathbb{P}\big(\gamma_k \geq \zeta_k \mid \gamma_k \geq \gamma_{th}\big) = 1$. Otherwise, if $\zeta_k > \gamma_{th}$, it follows that $\mathbb{P}\big(\gamma_k \geq \zeta_k \mid \gamma_k \geq \gamma_{th}\big) = \frac{\mathbb{P}\left(\gamma_k \geq \zeta_k, \gamma_k \geq \gamma_{th}\right)}{\mathbb{P}\left(\gamma_k \geq \gamma_{th}\right)} = \frac{\mathbb{P}\left(\gamma_k \geq \zeta_k\right)}{\mathbb{P}\left(\gamma_k \geq \gamma_{th}\right)}$. In the following Theorem, we provide a general expression of the conditional joint CCDF.

Theorem 4.2 *Conditional Joint CCDF: Given a MultiCell size of M, denote the set of the M tiers as $\mathcal{M}$. Let $\mathcal{M}_0$, $\mathcal{M}_1$, and $\mathcal{M}_2$ be the disjoint sets of the non-associated, associated and active, and associated but inactive tiers, respectively, such that $\mathcal{M}_0 \bigcup \mathcal{M}_1 \bigcup \mathcal{M}_2 = \mathcal{M}$. The conditional joint CCDF of the achievable SIRs on $\mathcal{M}_1$, given that the states of the remaining tiers $\mathcal{M}_0$ and $\mathcal{M}_2$ are known, is given by*

$$\mathbb{P}\left(\bigcap_{k\in\mathcal{M}_1}(\gamma_k \geq \zeta_k) \;\middle|\; \bigcap_{k\in\mathcal{M}}\mathcal{A}_k\right) = \frac{\mathbb{P}\left(\bigcap\limits_{k\in\mathcal{M}_1'}(\gamma_k \geq \zeta_k),\; \bigcap\limits_{k\in\mathcal{M}_1\cup\mathcal{M}_2\setminus\mathcal{M}_1'}\mathcal{A}_k\right)}{\mathbb{P}\left(\bigcap\limits_{k\in\mathcal{M}_1\cup\mathcal{M}_2}\mathcal{A}_k\right)}, \tag{4.21}$$

where the joint CCDF of the SIRs on different tiers is given in (4.19) and $\mathcal{M}_1'$ is the set of tiers for which $\zeta_k > \gamma_{th}$ and $\mathcal{A}_k$, $\forall k \in \mathcal{M}$ is defined as

$$\mathcal{A}_k := \begin{cases} \gamma_k \geq 0 & ; k \in \mathcal{M}_0 \\ \gamma_k \geq \gamma_{th} & ; k \in \mathcal{M}_1 \\ \gamma_k < \gamma_{th} & ; k \in \mathcal{M}_2 \end{cases}. \tag{4.22}$$

Proof: See Appendix B.2. ■

Corollary 4.1 *For a fixed $\zeta_k = \zeta$, the conditional joint CCDF of the achievable SIRs on $\mathcal{M}_1$ is*

$$\mathbb{P}\left(\bigcap_{k\in\mathcal{M}_1}(\gamma_k \geq \zeta) \;\middle|\; \bigcap_{k\in\mathcal{M}}\mathcal{A}_k\right) = \begin{cases} \frac{\mathbb{P}\left(\bigcap\limits_{k\in\mathcal{M}_1}(\gamma_k\geq\zeta),\; \bigcap\limits_{k\in\mathcal{M}_2}\mathcal{A}_k\right)}{\mathbb{P}\left(\bigcap\limits_{k\in\mathcal{M}_1\cup\mathcal{M}_2}\mathcal{A}_k\right)} & ; \zeta \geq \gamma_{th} \\ 1 & ; \zeta < \gamma_{th} \end{cases}, \tag{4.23}$$

Proof: The proof follows directly from Theorem 4.2. ■

4.3.4 Partial Offloading and Equal Task Division Approach

For any HTCU and for a given *MultiCell* size M, each link of a *Tier-k*, $k \in \mathcal{M}$ satisfies one of three states, namely, non-associated, associated-and-active, or associated-but-inactive, which yields 3^M different state combinations. Out of these 3^M combinations, there are 2^M combinations corresponding to a scenario where no offloading occurs and the whole task is processed locally. In particular, when all links of the M tiers are non-associated or associated but inactive. Let us denote $\mathcal{X}_M$ as the set of all possible state combinations for a given M such that the number of elements in $\mathcal{X}_M$ is $|\mathcal{X}_M| = X = 3^M$. Then, let $x_i^M \in \mathcal{X}_M$, $i \in \{1, 2, ..., 2^M, 2^M + 1, ..., 3^M\}$ be a specific state combination and define $\mathcal{M}_0^{x_i^M}$, $\mathcal{M}_1^{x_i^M}$, and $\mathcal{M}_2^{x_i^M}$ as the disjoint sets of the non-associated, associated and active, and associated but inactive tiers, respectively, corresponding to the state combination x_i^M. Hence, we can define

$$\omega^{x_i^M} = 1 + \left|\mathcal{M}_1^{x_i^M}\right| \tag{4.24}$$

as the number of task divisions corresponding to a certain combination $x_i^M \in \mathcal{X}_M$. Consequently, the average total delay experienced by the typical HTCU for a given M can be expressed as

$$\bar{T}_{h,s}^M = \mathbb{E}_{\mathcal{X}_M}(T_{h,s}^M) = \sum_{i=1}^{X} \mathbb{P}(x_i^M)\mathbb{E}[t_{h,s}^{x_i^M}], \tag{4.25}$$

where s refers to the adopted task division approach, *i.e.,* $s \in \{eq, adp\}$ denoting either equal or adaptive task division, $\mathbb{P}(x_i^M)$ is the probability of the combination x_i^M and $t_{h,s}^{x_i^M}$ is its corresponding processing delay given as follows:

$$\mathbb{P}(x_i^M) = \prod_{k \in \mathcal{M}_0^{x_i^M}} (1 - p_k) \times \prod_{k \in \mathcal{M}_{1\cup 2}^{x_i^M}} p_k \times \mathbb{P}\left[\bigcap_{k \in \mathcal{M}_{1\cup 2}^{x_i^M}} \mathcal{A}_k^{x_i^M}\right], \tag{4.26}$$

where p_k is given in (4.8), $\mathcal{M}_{1\cup 2}^{x_i^M} = \mathcal{M}_1^{x_i^M} \cup \mathcal{M}_2^{x_i^M}$, and the last term can be obtained from the joint link activation probability given in Theorem 4.1, and

$$t_h^{x_i^M} = \max\left(t_0^{x_i^M}, \max_{k\in\mathcal{M}_1^{x_i^M}} t_k^{x_i^M}\right), \tag{4.27}$$

where $t_0^{x_i^M} = \alpha_0^{x_i^M} t_{loc}$ is the delay of partial processing of part $\alpha_0^{x_i^M}$ of the task at the local processor and $t_k^{x_i^M} = \alpha_k^{x_i^M}(t_{k,tran}^{x_i^M} + t_{k,proc}^{x_i^M})$, $k \in \mathcal{M}_1^{x_i^M}$, previously given in (4.3), is the sum of the partial transmission and processing delays corresponding to offloading part $\alpha_k^{x_i^M}$ of the task to *Tier*-k for partial processing at the ECS therein. The final step is to derive an expression for the average processing delay $\mathbb{E}[t_{h,s}^{x_i^M}]$ corresponding to a certain state combination $x_i^M \in \mathcal{X}_M$ for each of the task division approaches. However, prior to this final step, we first consider some special cases of M for further illustration of the above discussion.

For example, $M = 0$ indicates the scenario where no offloading occurs and the whole task will be processed at the local processor. In such a scenario, we have only one combination x_1^0 and the total processing delay for the typical HTCU would be $t_{h,s}^{x_1^0} = t_{loc}$ regardless of the densities of HTCUs and SCs or the task division approach $s \in \{eq, adp\}$. For $M = 1$, we first enumerate the $X = 3^1$ different combinations as x_1^1 corresponding to the scenario where *Tier*-1 is non-associated, x_2^1 where *Tier*-1 is associated but inactive, and finally x_3^1 reflecting a scenario where *Tier*-1 is associated and active. For both x_1^1 and x_2^1 (2^1 combinations), our system reduces to the scenario where no offloading occurs such that $t_h^{x_1^1} = t_h^{x_2^1} = t_{loc} = C_u/f_{loc}$ and the probability sum of these two combinations is $\mathbb{P}(x_1^1) + \mathbb{P}(x_2^1) = (1 - p_1) + p_1\ \mathbb{P}(\gamma_1 < \gamma_{th})$ where p_1 is calculated from (4.8) and $\mathbb{P}(\gamma_1 < \gamma_{th}) = 1 - \mathbb{P}(\gamma_1 \geq \gamma_{th})$ is calculated from (4.11). For the third combination x_3^1, the computation task will be divided into $\omega^{x_3^1} = 2$ parts; one part $\alpha_0^{x_3^1}$ to be processed locally, and the remaining part $\alpha_1^{x_3^1}$ to be offloaded and processed at the ECS integrated within the SC of *Tier*-1. In this case, $\mathbb{P}(x_3^1) = p_1\ \mathbb{P}(\gamma_1 \geq \gamma_{th})$ and the processing delay

experienced by the typical HTCU corresponding to this specific combination is given by

$$t_h^{x_3^1} = \max_{k=0,1} t_k^{x_3^1} = max\left(\alpha_0^{x_3^1} t_{loc}\ ,\ \alpha_1^{x_3^1}\left(\frac{B_u}{R_1^{x_3^1}} + t_{ecs}\right)\right), \tag{4.28}$$

where $k = 0$ corresponds to the local processor and $R_1^{x_3^1}$ is the achievable rate on the link between the typical user and its closest SC given in (4.4). Then, the average delay experienced by the typical HTCU corresponding to the combination x_3^1 can only be calculated based on how the task is to be divided between the local processor and the ECS of *Tier*-1. In the following theorem, we calculate the average end-to-end processing delay under an equal task division approach. Then, in the following subsection, we introduce the proposed offline adaptive task division and calculate its corresponding average processing delay.

Theorem 4.3 *Assuming an SC can serve at most one HTCU, let $\mathcal{X}_M$ denote the set of all possible combinations in terms of tiers' association and activation of the typical HTCU for a given MultiCell size M. For a specific combination $x_i^M \in \mathcal{X}_M$, let $\mathcal{M}_0^{x_i^M}$, $\mathcal{M}_1^{x_i^M}$, and $\mathcal{M}_2^{x_i^M}$ be the disjoint sets of the non-associated, associated and active, and associated but inactive tiers, respectively. Then, the average end-to-end task processing delay experienced by the typical HTCU under an equal task division approach is given by (4.25) where*

$$\mathbb{E}[t_{h,eq}^{x_i^M}] = \frac{t_{loc}}{\omega^{x_i^M}} + \frac{G_M}{\omega^{x_i^M}} \times \int_{\min(\zeta_0,\gamma_{th})}^{\zeta_0} \frac{\left(1 - \mathbb{P}[\bigcap_{k\in\mathcal{M}_1^{x_i^M}} (\gamma_k \geq \zeta), \bigcap_{k\in\mathcal{M}_2^{x_i^M}} \mathcal{A}_k^{x_i^M}] / \mathbb{P}[\bigcap_{k\in\mathcal{M}_{1\cup 2}^{x_i^M}} \mathcal{A}_k^{x_i^M}]\right)}{(1+\zeta)\left[\log_2(1+\zeta)\right]^2} \mathrm{d}\zeta, \tag{4.29}$$

and the joint CCDF, t_{loc}, and $\omega^{x_i^M}$ are given in (4.19), (4.1), and (4.24), respectively. $G_M = \frac{MB_u}{W\ln(2)}$, $\zeta_0 = 2^{\frac{G_M \ln(2)}{t_{loc} - t_{ecs}}} - 1$, $\mathcal{M}_{1\cup 2}^{x_i^M} = \mathcal{M}_1^{x_i^M} \cup \mathcal{M}_2^{x_i^M}$, and $\mathcal{A}_k^{x_i^M}$ can be obtained from

(4.22) for a given x_i^M.

Proof: See Appendix B.2. ■

4.3.5 Adaptive Task Division

In this subsection, we propose a novel offline task division approach to enhance the delay performance compared to equal task division. In the proposed technique, using tools from stochastic geometry, we calculate adaptive α_k fractions that depend on the average performance of each *Tier-k* $\in \mathcal{M}$. It is noteworthy that similar to equal α_k, whenever a *Tier-k* is non-associated or associated but inactive, the corresponding α_k will be set to zero. However, in the proposed adaptive approach, the values of α_k will be predetermined for a given combination $x_i^M \in \mathcal{X}_M$ based on the average relative distances between the HTCUs and its serving SCs. It is noteworthy that optimizing the task division online based on the actual distances and channels between each user and its serving cells would force significant overhead and optimization delay which would affect the achievable gains and is out of the scope of this work. In the proposed adaptive task division approach, we aim at partitioning the task between the local processor and connected tiers $\mathcal{M}_1$ inversely proportional to their average expected individual delays of processing the whole task. Hence, we start by finding these individual delays, then, we calculate the corresponding task fractions in order to have on-average equal delays across all in-parallel processing resources.

Lemma 4.3 *Assuming the typical HTCU offloading the whole computation-intensive task to Tier-k, which is known to be associated and active i.e., $k \in \mathcal{M}_1$, while all other HTCUs follow multiple-association to reflect actual interference scenario, then, Tier-k individual*

delay is

$$T_k^{ind} = \begin{cases} t_{loc} & ;k=0 \\ t_{ecs} + G_M \times \int_{\gamma_{th}}^{\infty} \left(1 - \frac{\mathbb{P}[\gamma_k \geq \zeta]}{\mathbb{P}[\gamma_k \geq \gamma_{th}]}\right) \frac{\mathrm{d}\zeta}{(1+\zeta)[\log_2(1+\zeta)]^2} & ;k \in \mathcal{M}_1 \end{cases} \tag{4.30}$$

where $k=0$ denotes processing the whole task at the local processor, t_{loc} and t_{ecs} are given in (4.1) and (4.2), respectively, and $G_M = \frac{MB_u}{W\ln(2)}$.

Proof: See Appendix B.2. ■

Lemma 4.4 *Given a link state combination $x_i^M \in \mathcal{X}_M$, the fraction $\alpha_k^{x_i^M}$ of the computation-intensive task that should be allocated to Tier-k, $k \in \{0\} \cup \mathcal{M}_1^{x_i^M}$ is*

$$\alpha_k^{x_i^M} = \begin{cases} \frac{\prod_{i \in \{0\} \cup \mathcal{M}_1^{x_i^M}, i \neq k} T_i^{ind}}{\sum_{j \in \{0\} \cup \mathcal{M}_1^{x_i^M}} \left(\prod_{i \in \{0\} \cup \mathcal{M}_1^{x_i^M}, i \neq j} T_i^{ind} \right)} & ;k \in \{0\} \cup \mathcal{M}_1^{x_i^M} \\ 0 & ;\textit{otherwise} \end{cases}, \tag{4.31}$$

where T_i^{ind} is given in (4.30) and $\mathcal{M}_1^{x_i^M}$ denote the set of associated and active tiers for a given combination $x_i^M \in \mathcal{X}_M$.

Proof: Since the total delay is the maximum of the delays experienced over all tiers and the local processor, we allocate α_k such that the delays are on average equal to each other such that

$$\alpha_0 T_0^{ind} = ... = \alpha_k T_k^{ind} = ... = \alpha_{\left|\mathcal{M}_1^{x_i^M}\right|} T^{ind}_{\left|\mathcal{M}_1^{x_i^M}\right|} \qquad ;\forall k \in \mathcal{M}_1. \tag{4.32}$$

Solving the set of equations in (4.32) simultaneously along with $\sum_{k=0}^{\left|\mathcal{M}_1^{x_i^M}\right|} \alpha_k = 1$, and after some mathematical manipulations, the values of α_k, $k \in \{0\} \cup \mathcal{M}_1$ can be obtained in (4.31). ■

Theorem 4.4 *Following Theorem 4.3 and Lemma 4.4, without loss of generality, let*

$\alpha_{1'}^{x_i^M} \geq \alpha_{2'}^{x_i^M} \geq ... \geq \alpha_{M_1'}^{x_i^M}$, $\forall 1', 2', ..., M_1' \in \mathcal{M}_1^{x_i^M}$ *where* $M_1' = \left|\mathcal{M}_1^{x_i^M}\right|$. *Then, the average end-to-end task processing delay experienced by the typical HTCU under the adaptive task division approach is given by (4.25) where* $\mathbb{E}(t_{h,adp}^{x_i^M})$ *is given in (4.33) and*

$$\mathbb{E}(t_{h,adp}^{x_i^M}) = \frac{1}{\zeta_0^{x_i^M}} + \sum_{k'=1'}^{k'=M_1'} \int_{\min(\zeta_{k'}^{x_i^M}, \zeta_0^{x_i^M})}^{\min(\zeta_{k'+1}^{x_i^M}, \zeta_0^{x_i^M})} \left(1 - \frac{\mathbb{P}\left[\bigcap_{k=1'}^{k=k'} \gamma_k \geq \mathcal{Y}_k^{x_i^M}(y), \bigcap_{k \in \mathcal{M}_{1\cup 2}^{x_i^M} \setminus \{1', ..., k'\}} \mathcal{A}_k^{x_i^M} \right]}{\mathbb{P}\left[\bigcap_{k \in \mathcal{M}_{1\cup 2}^{x_i^M}} \mathcal{A}_k^{x_i^M} \right]} \right) \frac{\mathrm{d}y}{y^2}, \tag{4.33}$$

$$\zeta_k^{x_i^M} = \begin{cases} \frac{1}{\alpha_0^{x_i^M} t_{loc}} & ; k = 0 \\ \frac{\log_2(1+\gamma_{th})}{\alpha_k^{x_i^M}\left[\frac{MB_u}{W} + t_{ecs}\log_2(1+\gamma_{th})\right]} & ; k \in \mathcal{M}_1^{x_i^M} \\ \infty & ; k = M_1' + 1 \end{cases}, \tag{4.34}$$

$$\mathcal{Y}_k^{x_i^M}(y) = 2^{\left(\frac{(MB_u/W).y}{1/\alpha_k^{x_i^M} - t_{ecs}.y}\right)} - 1; \qquad k \in \mathcal{M}_1^{x_i^M}, \tag{4.35}$$

where $\mathcal{A}_k^{x_i^M}$ *is given in (4.22) and* $\mathcal{M}_1^{x_i^M}$ *denote the set of the associated and active tiers for a given* $x_i^M \in \mathcal{X}_M$. *Also,* $\mathcal{M}_{1\cup 2}^{x_i^M}$ *denote the set of all associated tiers regardless of link activation state.*

Proof: See Appendix B.2. ■

4.4 Simulation Results

In this section, we show the performance of the system considering the effect of different system parameters. In particular, we verify the accuracy of the derived mathematical analysis by carrying out Monte-Carlo simulations and comparing the obtained results with the derived analytical expressions.

4.4.1 Simulation Setup

We consider a simulation area of one km^2 with $1,000$ spatial realizations each of 10 different realization of the channels. We first distribute the HTCUs and SCs randomly and uniformly in the considered area according to their corresponding densities. Each HTCU attempts to associate with its M closest SCs. Each tier is allocated a portion $1/M$ of the available bandwidth, hence, no mutual interference exists among the different tiers. Whenever more than one HTCU is competing for the same SC through different tier orders, only one HTCU from the lower order tier will be served by the SC. If more than one HTCU is competing on the same tier, the served user is selected randomly by the SC. From the associated links, those which do not satisfy the minimum required SINR γ_{th} are dropped. Afterward, each HTCU splits its computation-intensive task according to the adopted division technique (equal or adaptive) among the local processor and the connected tiers.

The processing delays of all HTCUs are then averaged over several spatial and channel realizations. By comparing the obtained average delay with the derived analytical expressions in the previous section, we make sure that the obtained analytical formulae can be used later to study the performance of the system under different configurations easily and efficiently. Unless mentioned otherwise, the set of system parameters adopted in our simulations are as follows. The path loss exponent η is set to 4 to reflect an urban environment. The density of HTCUs λ_h is set to 500 HTCUs/km^2. For a UDN environment, the density of SCs and the co-located ECSs is taken as $5,000$ SCs/km^2. The order of the multiple-association M ranges from $M = 1$ and up to $M = 5$. The available bandwidth is equal to 180 kHz, however, only a fraction $\theta = 0.5$ of this bandwidth is allocated for the intensive-computation task offloading while the remaining bandwidth is left for the communication purposes, *i.e.,* $W = 0.5 \times 180$ kHz. The noise power density is -174 dBm/Hz [93]. The processing capacities of local and ECS processors are set to $f_{loc} = 100$ MHz and $f_{ecs} = 1$ GHz, respectively. The required computational capacity of

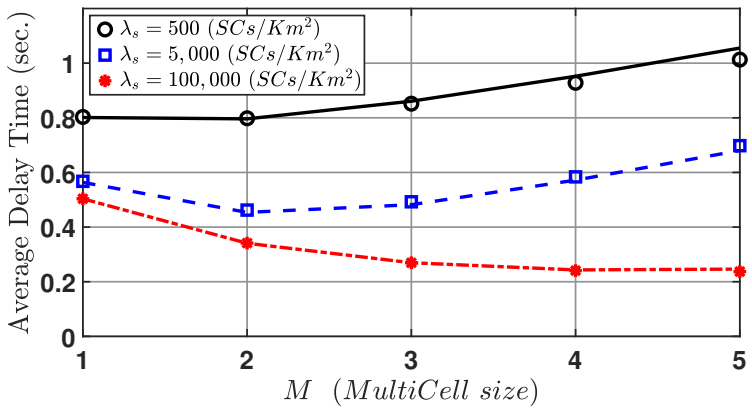


Figure 4.2: Average delay per user versus M under equal task division for different densities of small cells, $\gamma_{th} = 0$ dB, and $B_u = 10^5$ bits. Lines for analysis and markers for simulations.

the offloaded task is $C_u = 10^8$ CPU cycles which corresponds to $t_{loc} = 1$ second.

4.4.2 Equal Task Division

We begin with the equal task division approach (Equal-α) in which the task is divided equally among the local processor and the connected (associated and active) tiers. Fig. 4.2 shows the possible achievable gains from multiple-association when the density of SCs is much higher than that of HTCUs, which is the main characteristic of a UDN environment. When $\lambda_s \gg \lambda_h$, one can see that increasing the *MultiCell* size M significantly reduces the average overall delay of the HTCUs. However, when the density of the SCs is in the same order as the density of the HTCUs, the traditional single association where each HTCU associates with single SC results in the lowest delay. In other words, increasing the *MultiCell* size M causes a degradation in the performance since the low-order tiers (closer to the HTCUs) will be allocated smaller portions of the bandwidth while the high-order tiers (further away from the HTCUs) will mostly not be utilized due to the high competition among the HTCUs for the limited number of SCs.

In general, there exists an optimal value for the *MultiCell* size M which depends on several system parameters including the relative densities of HTCUs and SCs. For example, the middle curve in Fig. 4.2 shows that for a ratio of $\lambda_s/\lambda_h = 10$, $\gamma_{th} = 0$ dB,

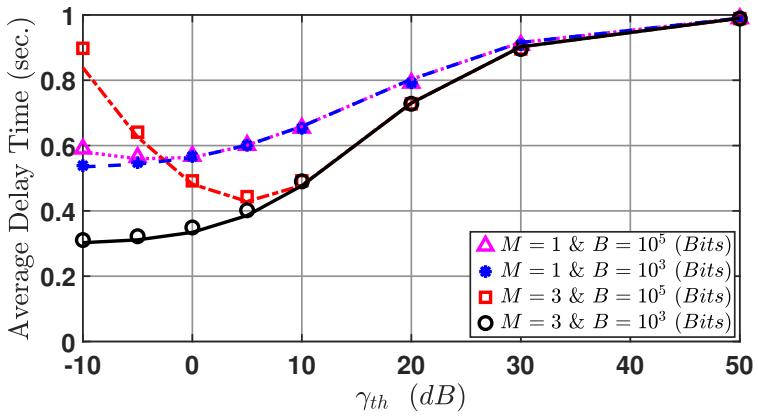


Figure 4.3: Average delay per user versus γ_{th} under equal task division for different M and B_u. Lines indicate analysis while markers represent simulations.

$B_u = 10^5$ bits, and $f_{ecs} = 1$ GHz, the minimum average delay can be achieved with double associations ($M = 2$) where the task is divided equally among the local processor, *Tier*-1, and *Tier*-2. On one hand, for a smaller value of M, *i.e.,* $M = 1$, the whole bandwidth is allocated to one tier which should reduce the transmission delay over that tier. However, the task therein is divided equally between only the local processor and *Tier*-1 which increases the task fraction allocated to each of them and yields higher processing times. On the other hand, although increasing the *MultiCell* size M offers smaller processing times on all tiers as well as the local processor, the transmission delays will dominate the total processing delay since less bandwidth will be allocated to each tier. Besides, the high-order tiers (further away from SCs) will suffer from poor channel conditions. Hence, either their transmission times will increase or even some of their links will drop which takes us back to the scenario of smaller M but with smaller bandwidth allocated to each tier. As the density of SCs (λ_s) increases, the optimal value of M gets larger since more tiers become closer to the HTCUs which enhances the qualities of their links. In return, more high-order tiers remain active and their transmission times get smaller.

In Fig. 4.3, we investigate the effect of the SINR threshold value γ_{th}. We show how a careful choice of γ_{th} given certain *MultiCell* size M and transmit-task size B_u significantly enhances the achievable delay performance. For example, one can notice

that for a relatively large transmit-task size $B_u = 10^5$ (bits) and $M = 3$, there exists an optimal value of γ_{th} around 5 dB which minimizes the average delay for given available bandwidth W and densities of SCs and HTCUs. On one hand, when γ_{th} is very small, this reflects a scenario where mostly all three tiers will be active regardless of the conditions of their links. Hence, offloading the relatively large portions of the tasks through these links, especially of the high-order tiers, will consume more time and increases the experienced delay. On the other hand, when γ_{th} is very large, this will deactivate all tiers and the whole task will be computed locally. The optimal value of γ_{th} will vary for different sets of system parameters, which motivates the existence of an analytical expression that can be accurately used to evaluate the performance of the system.

In addition, Fig. 4.3 also shows that for relatively small task sizes, decreasing γ_{th} yields lower average delay as the transmission times for offloading the relatively small portions of the tasks would be relatively small compared to the local processing time. Hence, allowing more tiers to be active and utilized for offloading reduces the portion of the tasks computed at the local processor. It is also clear from Fig. 4.3 that increasing the *MultiCell* size M yields further improvement in the achievable average delay, however, with the proper selection of the value of γ_{th}.

4.4.3 Adaptive Task Division

In this subsection, we show the additional gains that can be obtained by applying the proposed adaptive task division approach. In Table-4.1 above, we use (4.25) and (4.33) to show how the average delay per user can be optimized by the proper selections of the *MultiCell* size M for each set of different system parameters. Fig. 4.4 compares the average delay experienced when adopting Equal-α versus Adaptive-α for a fixed *MultiCell* size $M = 3$. One can notice a significant gain when the transmit-task size B_u is relatively small. Such behavior stems from the bottleneck delay being forced by the local processor when the Equal-α approach is adopted. In contrast, when the proposed Adaptive-α is

Table 4.1: Average delay per user versus *MultiCell* size M under the proposed adaptive task division approach for different SC densities λ_s, SINR threshold γ_{th}, and transmit task size B.

System Parameters			M=1	M=2	M=3
λ_s (SCs/km^2)	γ_{th} (dB)	B (bits)			
$5,000$	0	10^5	**0.4179**	0.4382	0.4623
$10,000$	0	10^5	**0.3369**	0.3444	0.3757
$10,000$	5	10^5	0.3271	**0.3098**	0.3219
$10,000$	10	10^5	0.3632	0.3194	**0.3153**
$5,000$	-10	10^4	0.2005	0.1558	**0.1545**
$10,000$	-10	10^4	0.1546	**0.1128**	0.1129
$10,000$	0	10^4	0.1682	0.0987	**0.0851**
$10,000$	10	10^4	0.2726	0.1599	**0.1222**

applied, smaller fractions of the tasks are allocated to the local processor. On the other hand, when the transmit-task size is relatively large, the gain from the proposed adaptive task division is not significant since the transmission delay dominates the total delay which forces the HTCU to allocate a relatively large fraction of the task to the local processor to reduce the transmission delay. It is noteworthy that the proposed adaptive task division approach still acts as a lower bound for the equal task division in terms of the experienced delay. Also, it is clear that γ_{th} still needs to be optimized under the proposed Adaptive-α approach. In the same scope, Fig. 4.5 shows how the transmit-task size B_u can influence the delays under different values of γ_{th}. Improper selection of γ_{th} (as discussed for Fig. 4.4) can lead to a longer delay when offloading the task compared to processing the whole task at the local processor.

The matching between simulation and analysis can be noticed for both Equal-α and Adaptive-α approaches for different system parameters as can be noticed in Fig. 4.2 - Fig. 4.5. Hence, the derived analytical expressions allow us to investigate the performance of the system under different conditions easily and efficiently. For example, Fig. 4.6 shows the delay performance versus the transmit-task size B_u under both task division approaches and for different values of M. Intuitively, the delay increases with B_u. However, the absolute value of the average delay and the rate of its increment varies from one scenario to another. On one hand, under equal task division, higher M gives

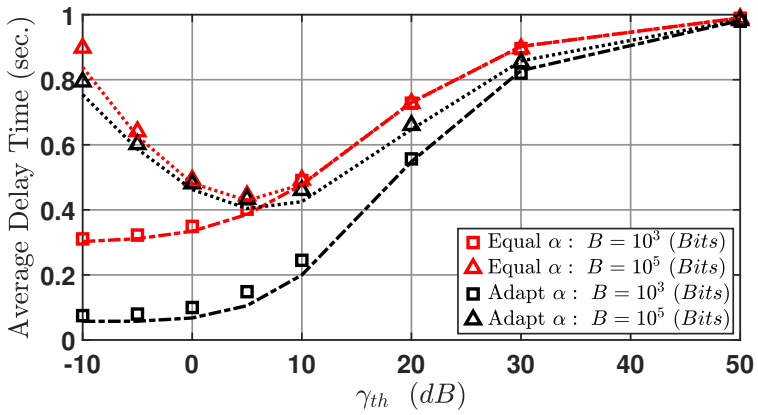


Figure 4.4: Average delay per user versus γ_{th} under both equal and adaptive task division and $M = 3$. Lines indicate analysis while markers represent simulations.

better performance when B_u is small (accordingly, transmission delay is small) thanks to allocating smaller parts of the processing task to local processing. However, the delay increases with B_u at much higher rate for larger M since the transmission delay of the high order tiers significantly increases with B_u. On the other hand, the rate at which the delay increases with B_u under adaptive task division remains within acceptable values thanks to the flexibility it offers. In particular, this results in from allocating smaller parts of the task to the high-order tiers, which have poor link conditions compared to the low-order ones. It is noteworthy that although the adaptive task division offers better performance than equal task division, the achievable gain starts to diminish with increasing the transmit-task size B_u.

From a different perspective, Fig. 4.7 shows the performance of the system against the ECS processing capacity f_{ecs} under fixed *MultiCell* size $M = 3$. For the equal task division approach, the task is divided into at most four equal parts with corresponding local processing delay of 0.25 seconds. However, some tiers may be disconnected due to competence among multiple HTCUs or due to degraded link quality. It can be noticed that when the transmit-task size is relatively small, no gain can be achieved from increasing f_{ecs} since the bottleneck of the experienced delay will come from the local processing delay. When the transmit-task size B_u increases, the transmission delay of offloading

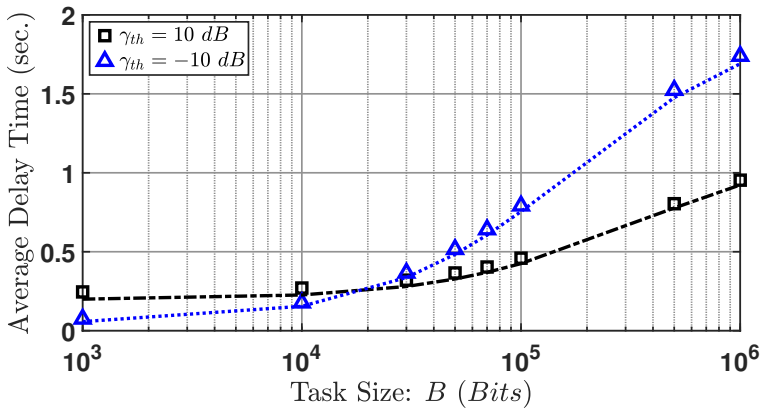


Figure 4.5: Average delay per user versus transmit-task size B_u in (bits) under adaptive task division, $M = 3$, and different γ_{th}. Lines indicate analysis while markers represent simulations.

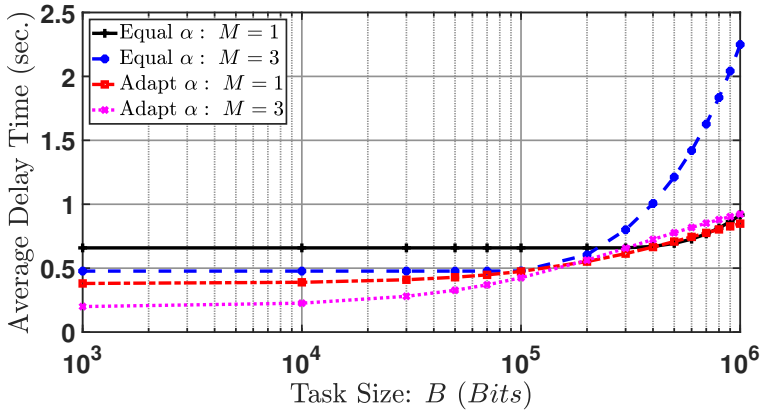


Figure 4.6: Average delay per user versus transmit-task size B_u in (bits) under equal and adaptive task division, different values of M, and $\gamma_{th} = 10$ (dB).

the task becomes the bottleneck for the system. Hence, we can initially observe some gain from increasing the processing capacities of the ECSs since the total delay comprises both transmission and processing delays. However, this gain diminishes swiftly with f_{ecs} when the processing delay of the ECS becomes incomparable to the transmission delay. On the other hand, for the adaptive task division approach, we can achieve significant gains from increasing f_{ecs} when the transmit-task size is relatively small. In other words, when the transmission delay is relatively small and the local delay can be also reduced by allocating smaller fraction α_0 of the task to the local processor, the overall delay can be significantly reduced by increasing f_{ecs}. Although for a larger transmit-task size there

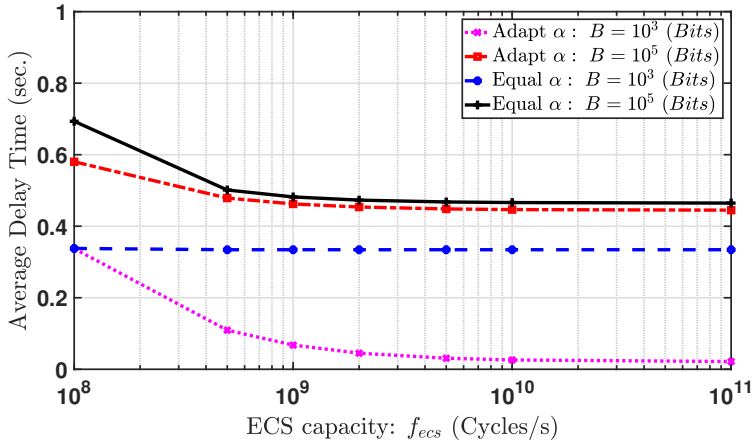


Figure 4.7: Average delay per user versus ECS processing capacity f_{ecs} $(cycles/sec.)$ under equal and adaptive task division, different B_u, $M = 3$, and $\gamma_{th} = 0$ (dB).

still exists some gain from applying the proposed adaptive task division technique, the achievable gain is not significant when the transmission delay dominates the total delay as discussed earlier. In particular, with higher values of f_{ecs}, the processing delay at the ECS becomes insignificant compared to the offloading transmission delay. Besides, the offloading transmission delay the and local processing delay become comparable, hence, no significant gain can be noticed from adapting the task division ratios.

4.5 Summary

We investigated the achievable gains from simultaneously offloading computation-intensive tasks of Human-Type Communication Users to multiple Small Cells equipped with Edge Computing Servers. In particular, we investigated the total delay experienced from the time instant of the task generation at the user until the whole task is fully processed. By doing so, we showed that the *MultiCell* size can be optimized for different system parameters including the densities of cells and users. In addition, we proposed an offline adaptive task division technique to further enhance the achievable gains from multiple-association compared to equal task division. We showed that an Ultra-Dense Network environment can exaggerate the achievable gains to support different use cases

targeted in 5G and beyond such as Ultra-Reliable Low-Latency Communications. Using tools from stochastic geometry, we formulated a mathematical framework and derived analytical expressions for the average experienced delay. Further extensions to this investigation could be optimizing resource allocation in terms of frequency, power, and MultiCell size as well as task division based on the instantaneous network conditions.

Chapter 5

Secrecy Performance in UDNs with Multiple-Association

5.1 Introduction

According to [2], the user rates would triple from 13.2 Mbps in 2018 up to 43.9 Mbps in 2023. With the frequent serious data breaches experienced recently, securing data has become more of a business risk rather than an information technology (IT) issue [2]. In this regard, UDN is seen as one key factor in future cellular networks where a UDN can provide the User Equipment UE with a virtual mobile network that follows its position [4]. The UE will always find close-proximity cells to activate, connect to, and be served. Physical Layer Security (PLS) is a promising approach to complement the upper layer cryptographic techniques in securing the enormous amount of data flying over the air [49]. Based on information-theoretic approaches, PLS exploits the inevitable distinction between the links of legitimate and illegitimate receivers of radio signals [48] to secure wireless data transfer.

In this chapter, we propose a secure multiple-association scheme in which the data traffic of a UE is split into different paths through the M closest cells. In addition to

mitigating the individual backhaul capacity limitations that can be found in UDNs [5]:

- we exploit the spatial diversity among the different paths to increase the achievable secrecy rate in the network.
- we derive a lower-bound analytical expression for the average secrecy per user using tools from stochastic geometry.
- We show that exploiting the proposed multiple-association in a UDN environment can mitigate the effect of several deteriorating factors such as the increasing densities of both UEs and eavesdroppers.

5.1.1 Related Works

Many recent investigations have been carried out in the context of PLS within a UDN environment [34, 94–97]. For example, the authors in [34] provided a novel mathematical framework for PLS in UDNs using tools from stochastic geometry. Considering a UDN environment, idle mode capabilities of the SCs and high probabilities of Line-of-Sight (LoS) communications were considered. In addition, the authors highlighted the performance in the case of both active and passive eavesdroppers which differ in the availability of their Channel State Information (CSI) at the transmitter. In [97], the authors proposed a new security metric called security pressure which is the maximum secured area provided during a legitimate connection. Then, they proposed a new user association policy in UDN to maximize the achievable security instead of the traditional association policy targeting Signal-to-Interference plus Noise Ratio (SINR) maximization. The authors in [97] considered a scenario of proactive eavesdroppers which are assumed to be intelligent nodes in choosing the active cells to attack. Those eavesdroppers are assumed to be non-uniformly distributed within a two-tiered heterogeneous UDN. In this regard, a secure user association scheme was proposed to maximize the network secrecy throughput.

Different from the above works, the authors in [94] considered PLS in the context of downlink Simultaneous Wireless Information and Power Transfer (SWIPT) scenario with both power splitting and time switching. The work presented therein was mainly focused on the mmWave frequency band in a UDN environment for its suitability for wireless power transfer. Using tools from stochastic geometry, the authors provided analytical expressions for the effective secrecy throughput in both cases of colluding and non-colluding eavesdroppers. The author also showed that proper selection of different system parameters such as power or time splitting ratio between data and power transfer, transmission power, and confidential information rate can be optimized the system perform acne in terms of security.

In [95], the authors proposed a secure user-centric clustering approach to serve the users by multiple cooperating small cells within a UDN environment. Two jamming strategies were suggested, namely, dedicated jamming in which a portion of the serving cells are dedicated to transmitting jamming signals, and embedded jamming in which the cells within the user-centric cluster jointly serve the user and produce jamming signals. The authors developed a set of heuristic greedy algorithms to maximize the secrecy performance with or without CSI of the eavesdroppers known at the serving cells. In a similar scenario in terms of cooperating SCs in UDN, the authors in [24] provided a novel mathematical framework for multiple-association in UDN environment, however, without consideration of secrecy performance in their system model.

The rest of the chapter is organized as follows. In section 5.2, we describe the system model. Section 5.3 provides the necessary mathematical derivations for the main and leakage link rates, respectively. The obtained results and discussion are reported in Section 5.4.

5.2 System Model

We consider a user-centric UDN where a massive number of deployed SCs are used to serve the existing UEs under the threat of coexisting illegitimate users aiming at eavesdropping the ongoing communications. SCs, UEs, and Eves are spatially distributed according to three independent HPPPs Φ_s, Φ_u, and Φ_e with corresponding densities λ_s, λ_u, and λ_e, respectively. Taking into consideration the UDN environment, hence, we assume that $\lambda_s \gg \lambda_u$ [4]. Following a practical scenario, the density of eavesdroppers is smaller than that of the legitimate users, *i.e.,* $\lambda_e \leq \lambda_u$ [34, 95]. In this investigation, we adopt a user-centric multiple-association scheme where each UE connects to and activates the closest M cells. The traffic of the served UE is split into a number of paths N_t, $1 \leq N_t \leq M$, depending on the availability of SCs. In doing so, we mitigate the limited backhaul capacities of the SCs and provide spatial diversity for the UEs in order to protect them from the eavesdropping of any illegitimate receiver. In other words, even if an eavesdropper has a relatively strong radio link with one of the serving SCs, the probability of having strong radio links with the other serving cells will decrease with the number of associated cells as illustrated in Fig. 5.1.

We assume that M orthogonal frequency bands are available in the network and that they are reused for each UE such that each SC exploits exactly one frequency band. In other words, each SC belongs to a specific tier where a *Tier-j* is the set of all SCs that are j^{th} closest to the UEs. For example, *Tier-1* is the set of closest SCs to the UEs. For the remaining SCs which do not belong to a specific tier, *i.e.,* they do not serve any UE, we assume idle mode capability in which the unused SC is switched-off to mitigate the ICI [4, 12]. Hence, the received signal which is transmitted from an SC on a specific tier suffers from Inter-Cell Interference (ICI) originating solely from those active cells on the same tier. In our system model, we take into consideration the different scenarios of an SC serving more than one UE. In such a scenario, when the different UEs associated with the same SC are served on the same tier, then, the allocated frequency band is

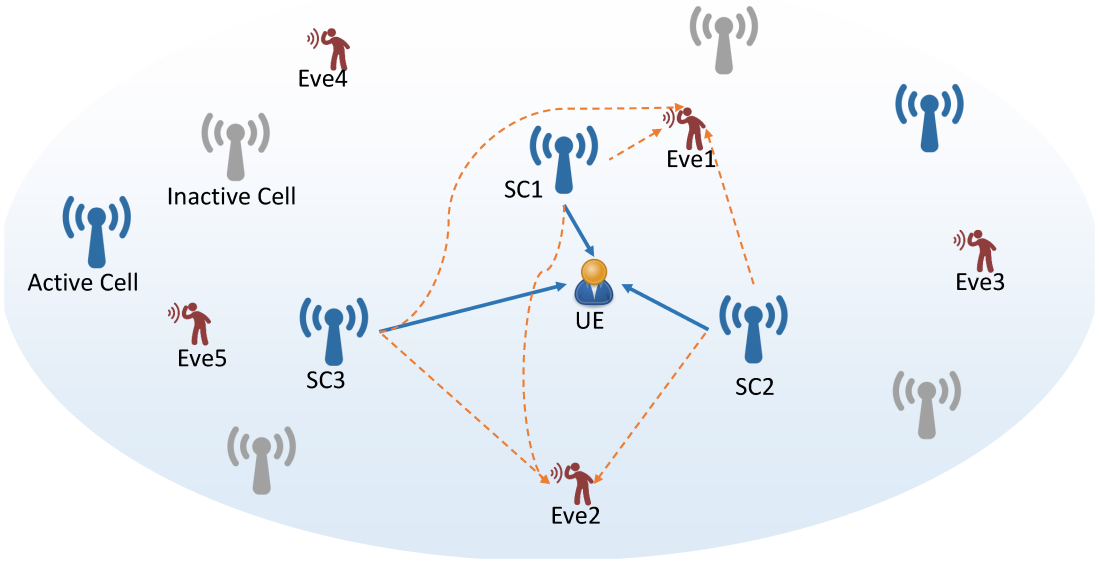


Figure 5.1: Multiple-associations scheme in a UDN with the presence of passive eavesdroppers.

divided equally and orthogonally in time or frequency among the different UEs. However, if they are associated with the same SC on different tiers, the higher-ordered tier (which corresponds to a further distance) is disconnected.

For the eavesdroppers, we assume non-colluding ones [98]. In such a case, each eavesdropper decodes the received from each serving SC independently from each other. To consider a worst-case scenario, we choose the one illegitimate receiver in the whole network which results in the minimum achievable secrecy-sum rate for the UE under investigation. To elaborate, we find the achievable secrecy rate from each tier separately for all eavesdroppers, sum over all tiers, then, find the minimum secrecy-sum and choose the corresponding Eve as the most detrimental eavesdropper (MDE).

We assume that all nodes are equipped with single omnidirectional antennas. The signal between a transmitter and a receiver attenuates according to $r^{-\eta}$ where r is the distance between transmitter and receiver and η is the standard path loss exponent with $\eta > 2$. In addition, we consider multi-path fading where Rayleigh fading channels are assumed with channel gains exponentially distributed with unity mean. Besides, a block

fading model is considered in which the channel gains are fixed over a Transmit Time Interval (TTI) and vary independently from one TTI to another. Finally, all SCs serving a specific UE are assumed to be infinitely backlogged with data ready to transmit.

5.3 Average Secrecy Rate per User

The objective of this section is to derive a lower bound for the average secrecy rate per user in a user-centric UDN with multiple-association. According to [35], the probability density function (PDF) of the distance between a typical UE located at the origin and the j^{th} closest SC r_j is

$$f_{r_j}(r) = \frac{2(\pi\lambda_s)^j}{(j-1)!} r^{2j-1} e^{-\pi\lambda_s r^2}, \tag{5.1}$$

where λ_s is the density of SCs. The idle mode probability in a UDN is given in [99]. Taking into consideration that a typical UE can activate up to j, $j \leq M$, SCs in the proposed multiple-association scheme instead of only one cell in single association, then, the idle mode probability reduces to

$$\mathbb{P}_{idle}^{j} = \left(1 + \frac{j\lambda_u}{3.5 + \lambda_s}\right)^{-3.5}, \tag{5.2}$$

with λ_u being the density of UEs. Hence, while keeping into consideration disconnecting high-ordered tiers in case of conflict with low-ordered tiers, one can interpret the probability of activation of SCs on a specific *Tier-j* as

$$\mathbb{P}_{a}^{j} = \left(\frac{3.5}{3.5 + \frac{(j-1)\lambda_u}{\lambda_s}}\right)^{3.5} - \left(\frac{3.5}{3.5 + \frac{j\lambda_u}{\lambda_s}}\right)^{3.5}. \tag{5.3}$$

Next, exploiting PLS to secure the communications of a typical UE on a specific

Table 5.1: Probability of Non-negative Secrecy

λ_e/λ_s	Tier-1	Tier-2	Tier-3	Tier-4	Tier-5
0.5	0.9781	0.9603	0.9449	0.9312	0.9181
0.25	0.9794	0.9644	0.9495	0.9380	0.9257
0.1667	0.9801	0.9655	0.9530	0.9409	0.9299
0.1	0.9813	0.9683	0.9560	0.9452	0.9354
0.0714	0.9823	0.9702	0.9590	0.9482	0.9374
0.05	0.9838	0.9718	0.9610	0.9501	0.9389

Tier-j, the achievable secrecy rate $\mathcal{R}_s^j$ can be given by [34]

$$\mathcal{R}_s^j = max\left(\mathcal{R}_m^j - \mathcal{R}_{e^*}^j,\ 0\right), \tag{5.4}$$

where $\mathcal{R}_m^j$ is the achievable rate on the main link for *Tier-j* and $\mathcal{R}_{e^*}^j$ is the leakage rate from *Tier-j* to the MDE, e^*. When $\lambda_s \gg \lambda_u, \lambda_e$ and $\lambda_u > \lambda_e$, then, the probability of $\mathcal{R}_m^j - \mathcal{R}_{e^*}^j > 0$ gets very high as shown in Table-5.1 for $\lambda_s = 5000$ SCs/km^2 and $\lambda_e = 50$ Eves/km^2. Hence, (5.4) can be simplified to

$$\mathcal{R}_s^j = \mathcal{R}_m^j - \mathcal{R}_{e^*}^j, \tag{5.5}$$

$$\mathbb{E}(\mathcal{R}_s^j) = \mathbb{E}(\mathcal{R}_m^j) - \mathbb{E}(\mathcal{R}_{e^*}^j), \tag{5.6}$$

where $\mathbb{E}[.]$ denotes expectation of a random variable.

Hence, we aim next to finding the average rates of both main and leakage links from each tier. For the expected value of the main link rate, it can be derived following the same steps in [23].

Theorem 5.1 *In a user-centric UDN with multiple-association, the average achievable rate per user on Tier-j, $1 \leq j \leq M$ is given by*

$$\mathbb{E}(\mathcal{R}_m^j) = B_j \frac{\mathbb{P}_a^j \lambda_s}{\lambda_u} \int_0^\infty \frac{1}{\left(\mathcal{A}_m^j(t)\right)^j} \mathrm{d}t, \tag{5.7}$$

$$\mathcal{A}_m^j(t) = 1 + \mathbb{P}_a^j (2^t - 1)^{\frac{2}{\eta}} \int_{(2^t-1)^{-\frac{2}{\eta}}}^{\infty} \frac{1}{1+u^{\frac{\eta}{2}}} \mathrm{d}u, \tag{5.8}$$

where B and $\mathbb{P}_a^j$ are the allocated bandwidth in Hz and the probability of active cells obtained from (5.3) on Tier-j, respectively.

Proof: The achievable rate by a typical user on *Tier-j* exploiting the total available bandwidth B_j is given by

$$R_m^j = B_j \log_2(1 + \frac{P_j h_j r_j^{-\eta}}{I_j + \sigma^2}), \tag{5.9}$$

where P_j is the transmission power of the serving cell on *Tier-j*, h_j, r_j are the channel gain and distance between the typical user and its serving cell on *Tier-j*, I_j is the ICI sensed at the typical user from all active cells on *Tier-j* and σ^2 is the power of the Additive White Gaussian Noise (AWGN) at the receiver. The fraction of UEs served on the total bandwidth of *Tier-j* is given by $\frac{\mathbb{P}_a^j \lambda_s}{\lambda_u}$. Hence, the average achievable rate per user on *Tier-j*

$$\mathbb{E}(\mathcal{R}_m^j) = \frac{\mathbb{P}_a^j \lambda_s}{\lambda_u} \mathbb{E}_{r_j, h_j, I_j}[R_j], \tag{5.10}$$

where

$$I_j = \sum_{k \in \Phi_j^a \backslash b_{j_0}} P_j h_k r_k^{-\eta}. \tag{5.11}$$

Φ_j^a is the thinned HPPP of all active cells on *Tier-j*, b_{j_0} is the serving cell, P_j is the transmission power of a cell on *Tier-j* which is assumed fixed, and h_k and r_k are the channel gains and distances between an interfering cell $k \in \Phi_j^a \backslash b_{j_0}$ and the typical UE. Following the same steps in [23],

$$\begin{aligned} \mathbb{E}_{r_j, h_j, I_j}[R_j] &= \int_0^{\infty} \mathbb{P}\left[R_j > t\right] \mathrm{d}t \\ &= \int_0^{\infty} \int_0^{\infty} e^{-(2^t-1)\frac{\sigma^2}{P} r^{\eta}} \mathcal{L}_{I_k'}((2^t-1)r^{\eta}; r) f_{r_k}(r) \, \mathrm{d}r \mathrm{d}t, \end{aligned} \tag{5.12}$$

and the Laplace transform of the normalized ICI $I'_j = \frac{I_j}{P_j}$ is given by

$$\mathcal{L}_{I'_j}(s;r) = \exp\left(-\pi\ \mathbb{P}_a^j \lambda_s\ s^{\frac{2}{\eta}} \int_{\frac{r^2}{s^{2/\eta}}}^{\infty} \frac{1}{1+u^{\frac{\eta}{2}}} \mathrm{d}u\right). \tag{5.13}$$

After some mathematical manipulations, (5.7) and (5.8) can be obtained, which completes the proof. ■

For the leakage link rate, we consider a worst-case scenario for the MDE in which we separately consider the closest eavesdropper to each serving cell as the MDE. By doing so, we guarantee that the obtained formula for the secrecy rate constitutes a lower bound for the actual achievable secrecy rate. The expected value of the leakage link rate from each tier is given in the following theorem.

Theorem 5.2 *In a user-centric UDN with multiple-association, the average maximum leakage rate per user on Tier-j, $1 \leq j \leq M$ is given by*

$$\mathbb{E}(\mathcal{R}_e^j) = B_j \frac{\mathbb{P}_a^j \lambda_s}{\lambda_u} \int_0^{\infty} \frac{1}{\mathcal{A}_e^j(t)} \mathrm{d}t, \tag{5.14}$$

$$\mathcal{A}_e^j(t) = 1 + \frac{\mathbb{P}_a^j \lambda_s}{\lambda_e} (2^t - 1)^{\frac{2}{\eta}} \int_{(2^t-1)^{-\frac{2}{\eta}}}^{\infty} \frac{1}{1+u^{\frac{\eta}{2}}} \mathrm{d}u, \tag{5.15}$$

where λ_e is the density of eavesdroppers and $\mathbb{P}_a^j$ is obtained from (5.3).

Proof: The most detrimental eavesdropper (MDE) is the one which causes the typical UE to achieve the minimum secrecy-sum rate from all tiers. However, for the sake of tractability, when deriving the achievable secrecy rate on a specific *Tier-j*, we consider the MDE as the closest eavesdropper to the serving cell on *Tier-j*, and changes from one tier to another. Hence, the PDF of the distance between the serving cell (regardless of its tier order) and the MDE will be given as

$$f_{r_e}(r) = 2\pi\lambda_e r e^{-\pi r^2}. \tag{5.16}$$

The achievable SINR of the MDE on *Tier-j* is given by

$$\gamma_e^j = \frac{P_j h_e r_e}{I_{j,e} + \sigma^2}, \tag{5.17}$$

where P_j is the transmission power, $h_e,\ r_e$ are the channel gain and distance between the serving cell on *Tier-j* and the MDE, $I_{j,e}$ is the ICI sensed at the MDE from all active cells on *Tier-j* except the one on which it eavesdrops, and σ^2 is the power of the AWGN. The Laplace transform of the PDF of the normalized ICI $I'_{j,e} = \frac{I_{j,e}}{P_j}$ can be given as

$$\begin{aligned}
\mathcal{L}_{I'_{j,e}}(s;r) &= \mathbb{E}_{I'_{j,e}}\left[e^{-sI'_{j,e}}\right] \\
&= \mathbb{E}_{\Phi_j^a, h_e}\left[\exp\left(-s\sum_{k\in\Phi_j^a\setminus b_{j_0}} h_e r_e^{-\eta}\right)\right] \\
&= \mathbb{E}_{\Phi_j^a}\left[\prod_{k\in\Phi_j^a\setminus b_{j_0}} \mathbb{E}_{h_e}\Big[\exp(-sh_e r_e^{-\eta})\Big]\right] \\
&= \exp\left(-\mathbb{P}_a^j \lambda_s \int_0^{2\pi}\int_r^{\infty} \left(1 - \mathbb{E}_{h_e}\left[e^{-sh_e r_e^{-\eta}}\right]\right) r_e \ \mathrm{d}r_e \ \mathrm{d}\theta\right) \\
&= \exp\left(-\pi \ \mathbb{P}_a^j \lambda_s \ s^{\frac{2}{\eta}} \int_{\frac{r^2}{s^{2/\eta}}}^{\infty} \frac{1}{1+u^{\frac{\eta}{2}}} \mathrm{d}u\right),
\end{aligned} \tag{5.18}$$

For the sake of tractability, $I_{j,e}$ is approximated by the interference sensed from those active cells on *Tier-j* while assuming that they are further than r_e^j. This will also give a higher bound on the leakage rate since the actual interferes may be closer than r_e^j. Similar

to the proof of the main link rates,

$$\begin{aligned}
\mathbb{P}[\gamma_e^j > \epsilon] = \; & \mathbb{E}_{I_{j,e},r_e}\left[\mathbb{P}\left[h_e^j > \epsilon(\frac{\sigma^2}{P_j} + I'_{j,e}){r_e}^{\eta}\right] \mid I'_{j,e}, r_e\right] \\
= \; & \mathbb{E}_{I'_{j,e},r_e}\left[\exp\Big(-\epsilon(\frac{\sigma^2}{P_j} + I'_{j,e}){r_e}^{\eta}\Big) \mid I'_{j,e}, r_e\right] \\
= \; & \int_{r>0} e^{-\epsilon\frac{\sigma^2}{P_j}r^{\eta}} \mathbb{E}_{I'_{j,e}}\left[\exp\Big(-\epsilon I'_{j,e}r^{\eta}\Big) \mid I'_{j,e}, r\right] f_{r_e}(r)\, \mathrm{d}r \\
= \; & \int_0^{\infty} e^{-\epsilon\frac{\sigma^2}{P_j}r^{\eta}} \mathcal{L}_{I'_{j,e}}(\epsilon r^{\eta}; r) f_{r_e}(r)\, \mathrm{d}r. \qquad (5.19)
\end{aligned}$$

Assuming an interference-limited scenario which is a common assumption in UDNs for any path loss exponent η [100], the average maximum leakage rate per user on *Tier-j* can be calculated as follows

$$\mathbb{E}(R_e^j) = B_j \int_0^{\infty} \mathbb{P}\left[\gamma_e^j > 2^t - 1\right] \mathrm{d}t. \qquad (5.20)$$

Keeping in consideration that leakage rate is proportional to the fraction of UEs being served on *Tier-j*, using (5.19) and performing some mathematical manipulations completes the proof. ■

5.4 Results and Discussion

In this section, the verification of the obtained analytical expressions is derived. The path loss exponent is set to $\eta = 4$ to reflect an urban environment [22], the allocated bandwidth per tier is fixed and equal to 180 kHz, transmission power per cell is set 23 dBm, and the thermal noise is set to -174 dBm/Hz. The small scale fading is assumed Rayleigh where the channel gains are exponentially Independently and Identically Distributed (i.i.d.) with unity mean. We sweep different densities of small cells, users, and eavesdroppers. Unless otherwise stated, the densities are set follows; the SCs' density $\lambda_s = 5,000$ SCs/km^2, the users' density $\lambda_u = 500$ users/km^2, and the eavesdroppers'

density $\lambda_e = 50$ Eves/km^2. The number of SCs serving a UE is set to $M = 1,\ 2$, and 3. The Monte-Carlo simulations are averaged over 500 spatial realization each with 100 different channel realization and over an area of one km^2.

Fig. 5.2 shows the total average secrecy rate per user from all associated tiers for different numbers of serving cells (*MultiCell* sizes). The obtained results show the significant gain achieved from the Ultra-Dense Network (UDN) environment where the achievable average secrecy rate per user increases with the small cell density λ_s. This gain stems from the shorter distances among the UE and its serving cells. In addition, when the serving cells get closer to the UE, this guarantees better protection against eavesdroppers which statistically get further from the cells. On the other hand, it is also clear the increasing the *MultiCell* size M boosts the achievable secrecy rate. Although a UE being served by the second or third (up to M^{th}) closest cell degrades the channel between the UE and its serving cells, it provides spatial diversity against the eavesdropper and helps mitigate any backhaul capacity limitations that can be found in UDNs [5]. It is to be noted that the achievable gains from network densification and multiple-association come at the cost of increased network resource usage such as the deployment of extra cells, energy consumption, and exploitation of additional frequency bands, respectively. However, this cost represents a compromise that must be made to tackle the unprecedented requirements in 5G networks and beyond in terms of user capacity and security.

In Fig. 5.3, we show that the total average secrecy rate per user from the M different associated cells decreases when the density of UEs λ_u increases. Since we consider a user-centric approach, it is important to study the average performance achievable by an individual user. In this regard, we show that the proposed multiple-association scheme with $M = 3$ can almost achieve the same performance as the traditional single association scheme, however, with a 10-fold increase in the density of supported UEs from 100 UEs/km^2 to 1000 UEs/km^2 under the same other system parameters.

In Fig. 5.4, one can notice a slight degradation in the achievable secrecy perfor-

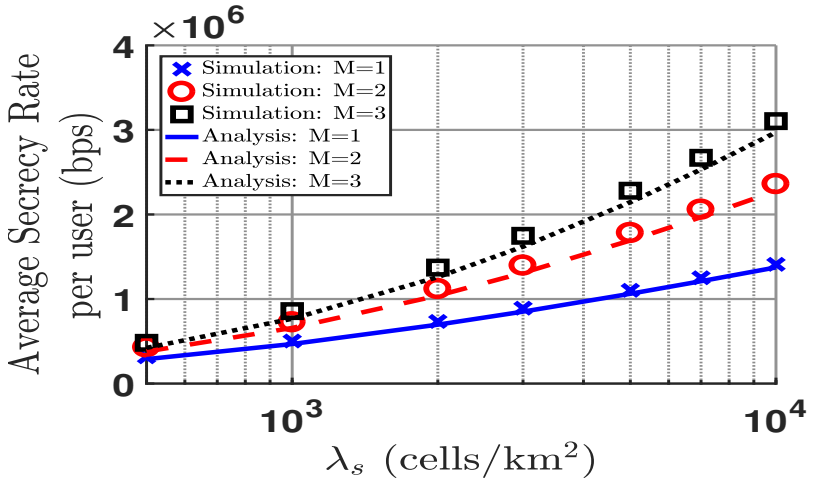


Figure 5.2: Average secrecy rate per user versus the small cell density λ_s for different *MultiCell* sizes M with density of users $\lambda_u = 500$ users/km^2 and relative density of Eves $\lambda_e/\lambda_u = 0.01$.

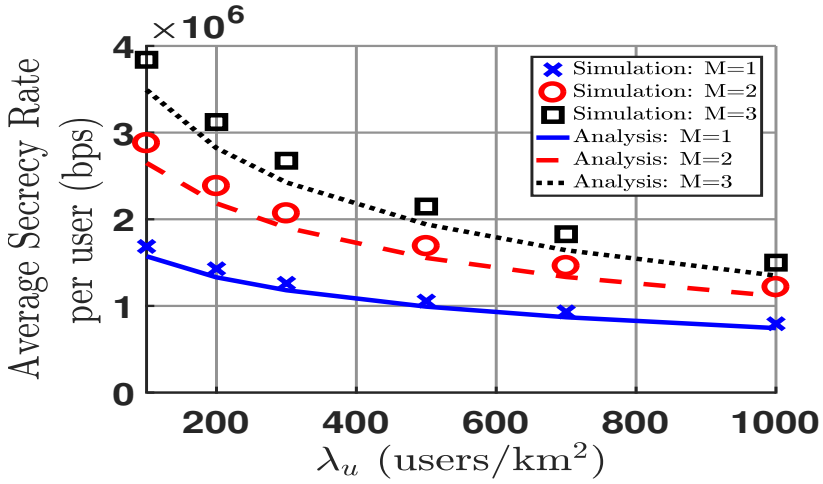


Figure 5.3: Average secrecy rate per user versus the density of users λ_u for different *MultiCell* sizes M with density of small cells $\lambda_s = 5,000$ cells/km^2 and density of Eves $\lambda_e = 50$ Eves/km^2.

mance when the density of eavesdroppers λ_e increases. However, this slight decrease due to 20-fold increase in λ_e can be fully mitigated by multiple-association with only $M = 2$ compared to a single association scheme.

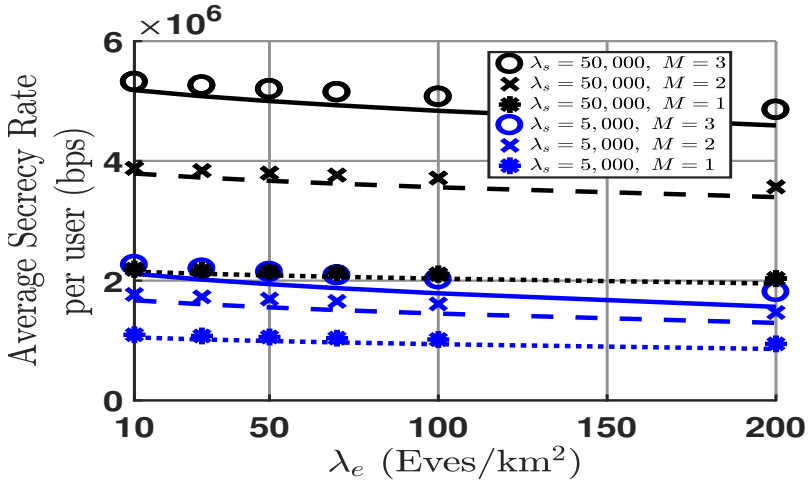


Figure 5.4: Average secrecy rate per user versus the density of eavesdroppers λ_e for different *MultiCell* sizes M and small cells densities λ_s with density of users $\lambda_u = 500$ users/km^2. Lines indicate analysis while markers represent simulations.

5.5 Summary

In this chapter, we propose a user-centric Physical Layer Security approach to increase the average secrecy rate per user using multiple associations of small cells to a User Equipment. In doing so, a typical UE connects to and activates the closest M cells where the data traffic of this UE is split into multiple paths ($\leq M$). This approach offers a double benefit for the network. First, it provides spatial diversity against the possible eavesdroppers whose locations are not identified by the network. Second, it helps mitigate the individual cell backhaul capacity limitations usually found in UDNs. A mathematical framework is formulated using tools from stochastic geometry to derive a lower bound for the average secrecy rate per user. The obtained results show how the secrecy performance is improved by the increasing number of serving cells M and/or density of small cells. On the other hand, the increasing densities of both users, and eavesdroppers result in degradation in the average secrecy rate per user which can be fully mitigated by increasing M. An important extension to this work would be the integration of Multiple-Input Multiple-Output (MIMO) systems to direct the signals to the targeted legitimate receiver.

Chapter 6

A Hybrid NOMA/OMA Scheme to Support mMTC in UDNs

6.1 Introduction

According to [2], amongst the different categories of mobile devices, Machine-to-Machine (M2M) connections are anticipated as the fastest-growing category with a 30 percent Compound Annual Growth Rate (CAGR) between 2018 and 2023. Consequently, tremendous advances in enabling technologies are needed in the next generations of cellular communications. Recently, Non-Orthogonal Multiple-Access (NOMA) has been considered extensively in academia and it was considered as a study item of 5G New Radio (NR) [44]. In NOMA [42], several users (two or more) share the same time-frequency-spatial resources simultaneously exploiting either the power domain or code domain.

In this chapter, we propose a hybrid NOMA/OMA scheme to support a massive number of MTCDs existing in a UDN environment. In particular:

- We divide the MTCDs into two groups based on their normalized Signal-to-Interference Ratio (SIR) and pair them to form NOMA clusters.
- We show that a SIR threshold value exists to maximize the performance gain of the

NOMA/OMA scheme compared to an only OMA-based scheme. However, densifying the network with more cells can compromise the gain obtained from NOMA.

- Analytical expressions for the Area Spectral Efficiency (ASE) gain are derived using tools from stochastic geometry and their accuracy is verified through simulation results

6.1.1 Related Works

Despite the potential of NOMA, the earned gains were not significant enough to continue with NOMA as a work item for the NR. It was decided to postpone NOMA for possible next generations of cellular networks to tackle the existing challenges [45]. Amongst the challenges obstructing the efficient deployment of NOMA is the NOMA clustering process [46]. In particular, which users will share the same resource, what is the optimal cluster size, and prior to these, what is the objective of a NOMA cluster, *i.e.,* maximizing the network throughput or achieving fairness among the different users? For example, the authors in [47] proposed a resource allocation scheme with a controlling parameter that governs the trade-off between the two objectives. While Successive Interference Cancellation (SIC) is responsible for mitigating the intra-cell or Inter-NOMA Interference (INI), a non-negligible complexity is added to the receiver.

Also, imperfect SIC significantly degrades the gain obtained from adopting NOMA [47]. In the same scope, ordering the users in the SIC process based on either the instantaneous received power or on the mean received power (distance) should be considered. For the former, *i.e.,* the instantaneous received power, Channel State Information (CSI) should be acquired which increases the overhead in the system. However, inaccurate or non-optimal performance may result from using the distance instead. The authors in [101] studied the probability that distance-based and instantaneous signal power-based rankings of NOMA users will yield the same results. In that, they investigated how this probability is affected by the different system parameters and its impact on the coverage

probability of both uplink and downlink.

Back to the clustering process, many existing works consider random clustering of users [68,101]. However, some other works proposed some sort of selective clustering based on either average received power (distance) [102] or instantaneous received power [68], with or without consideration of Inter-Cell Interference (ICI). It is noteworthy that clustering and ordering of users for SIC are two independent operations [46].

The rest of the chapter is organized as follows. In Section 6.2, we describe the system model. The analysis of the performance and the derived analytical expressions are obtained in Section 6.3. Section 6.4 presents the obtained numerical results and gained insights.

6.2 System Model

We consider a network of a massive number of Machine-Type Communication Devices (MTCDs) to be served in the downlink. Such a scenario lies under the use case of mMTC targeted in the 5G and beyond. We consider a UDN scenario where a massive number of Small Cells (SCs) are used to serve those MTCDs as shown in Fig. 6.1. Both SCs and MTCDs are spatially distributed according to two independent Homogeneous Poisson Point Processes (HPPPs), Φ_s and Φ_m with densities λ_s and λ_m, respectively. We assume an activity profile with active probability ρ_m for the MTCDs. As a result, we will have a thinned Homogeneous Poisson Point Process (HPPP) Φ_m^a with density $\lambda_m^a = \rho_m \lambda_m$ for the active MTCDs. In our model, we assume a heavily loaded regime such that $\lambda_m^a \gg \lambda_s$ which coincides with the mMTC scenario even in a UDN environment. Note that only active cells where at least one user is served by the SC are considered for both ICI and ASE calculations.

A hybrid NOMA/OMA system is deployed for the multiple access of the massive number of active MTCDs. In OMA, each user is allocated one Resource Block (RB) with

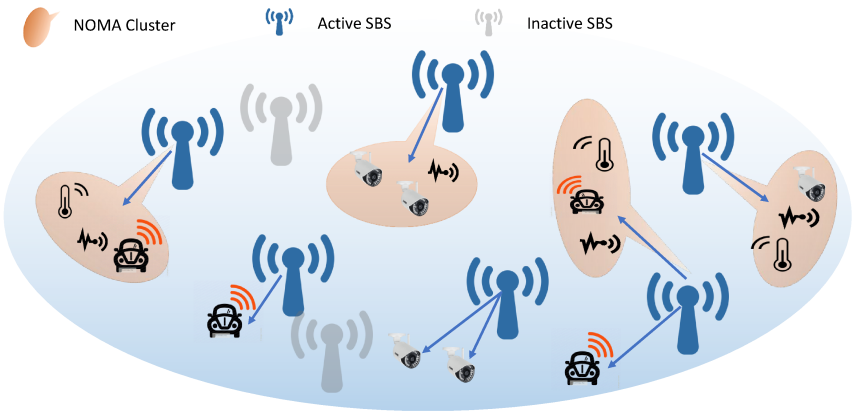


Figure 6.1: An illustration of the considered system model.

a normalized bandwidth of 1 Hz. However, in NOMA, multiple users sharing the same resources are allocated a number of RBs equal to the size of the NOMA cluster for a fair comparison. Without loss of generality, we consider a two-user NOMA scenario and divide the existing users into two complement disjoint groups with an SIR threshold value T separating them. Then, two randomly selected users, one from each group, are paired to form the NOMA cluster. This system model can be further extended by increasing the number of complement groups, hence, the size of the NOMA cluster. However, this would increase the complexity of the required SIC [103]. We use tools from stochastic geometry to model the performance of the network in terms of the ASE gain compared to pure OMA schemes.

In the considered scenario, all MTCDs and SCs are equipped with single omnidirectional antennas. The SEPL model is assumed for the large-scale fading, hence, the received signals at the devices are attenuated by a factor $e^{\alpha r^{\beta}}$ where r is the distance between a transmitter and receiver. Also, we assume Rayleigh small-scale fading channels with power channel gains h, such that $h \sim \exp(1)$ and channel state information (CSI) known at the transmitter [68]. Moreover, a full-buffer traffic model is considered in which active MTCDs always have data ready to transmit. We assume a frequency reuse factor of one and each cell has enough resources to support the connections of associated devices using a contention-free access protocol. Superposition coding is used to transmit the

message of two or more devices simultaneously via exploiting the power domain. At the receiver side, MTCDs perform SIC to decode their desired signals. Messages transmitted with lower powers are treated as interference when decoding a specific message. In the case of OMA, the achievable rate per user per unit bandwidth is given by

$$R_i^O = \log_2\Big(1 + \frac{Ph_i e^{-\alpha r_i^\beta}}{I_i + \sigma^2}\Big), \tag{6.1}$$

where P is the transmit power of the SC. h_i, r_i are the power gain and distance of the channel between the SC and device i and I_i is the ICI sensed at device i. α and β are the SEPL parameters and σ^2 is the Additive White Gaussian Noise (AWGN) power. For the devices adopting NOMA, a user k within a NOMA cluster of size K will receive its indicated message superposed with other messages of higher and/or lower allocated power levels. Hence, the achievable rate per unit bandwidth would be

$$R_k^N = \log_2\Big(1 + \frac{\varepsilon_k Ph_i e^{-\alpha r_i^\beta}}{\zeta I_{k,h}^N + I_{k,l}^N + I_i + \sigma^2}\Big). \tag{6.2}$$

ζ is the SIC error propagation factor, ε_k is the power fraction allocated to user k, and $I_{k,h}^N = \sum_{i=0}^{k-1} \varepsilon_i Ph_i e^{-\alpha r_i^\beta}$ is the INI due to messages with higher allocated powers. This interference should be ideally removed if $\zeta = 0$. However, imperfect SIC may cause residual interference reflected by $\zeta > 0$. $I_{k,l}^N = \sum_{i=k+1}^{K} \varepsilon_i Ph_i e^{-\alpha r_i^\beta}$ is the INI caused by messages with lower allocated powers and cannot be removed. The ASE can be then calculated as

$$ASE = \lambda_a \mathbb{E}[R_{cell}], \tag{6.3}$$

where λ_a is the density of active SCs and R_{cell} is the achievable rate per cell. It is noteworthy that assuming a typical MTCD at the origin which is allocated the whole bandwidth, the achievable rate by this device reflects the cell rate. However, one must take into consideration the different scenarios of this typical device. In our proposed hybrid NOMA/OMA scheme, the existing users are divided into two groups based on

their normalized SIR [68],

$$c_i = \frac{Ph_i e^{-\alpha r_i^{\beta}}}{I_i} = \frac{Ph_i e^{-\alpha r_i^{\beta}}}{\sum_{\Phi_s \setminus b_0} Pg_i e^{-\alpha v_i^{\beta}}}, \tag{6.4}$$

where h_i and r_i correspond to the channel between user i and its serving cell while g_i and v_i correspond to all other channels between user i and interfering cells. The first group *(Group1)* contains all the users with $c_i > T$ and the second group *(Group2)* contains all the users with $c_i < T$ where T is the normalized SIR threshold separating the two groups. In our system model, we consider all existing users in contrast to [46, 68, 102], which considers only those users adopting NOMA.

As we increase the SIR threshold T, the users in *Group1* achieve higher normalized SIR but the number of users significantly decreases. Hence, we show that an optimum point of T exists, at which the average achievable rate is maximized. This optimal point will depend on the different parameters of the system. However, the number of users adopting NOMA will always be maximized when the users are divided into two statistically equal groups *i.e.,* $T = C_0$; $F_c(C_0) = 0.5$ and $F_c(C)$ is the Cumulative Distribution Function (CDF) of the normalized SIR of the existing users.

To this end, we aim to find an analytical expression that reflects the performance of our proposed system model and show that an optimum point that maximizes the gain of this hybrid NOMA/OMA scheme exists.

6.3 Performance Analysis

Since both SCs and MTCDs are spatially distributed according to two independent HPPPs, we recall that the distribution of the number of users per cell (N) is random. The probability density function (pdf) of the size of a typical Voronoi cell was derived empirically using Monte-Carlo simulations in [66]. Then, this distribution was used in [67] to obtain an expression for the probability mass function (pmf) of the number of users

per cell N which depends solely on the average density of users per cell $\lambda = \frac{\lambda_u}{\lambda_s}$, where

$$f_N(n) = \frac{(\frac{3.5}{\lambda+3.5})^{3.5}}{\Gamma(3.5)} \left(\frac{\lambda}{\lambda+3.5}\right)^n \frac{\Gamma(n+3.5)}{\Gamma(n+1)}. \tag{6.5}$$

Note that $f_N(0)$ reflects the probability of an inactive cell referred to as idle mode capability in UDNs. In idle mode, empty cells are switched off to mitigate the ICI. However, when the number of devices increases as in the mMTC scenario, the probability of SCs switching off significantly decreases. Next, the normalized SIR distribution can be obtained by following the steps in [22, 104]:

$$\begin{aligned} F_c(C) &= \mathbb{P}[c < C] && (6.6)\\ &= 1 - \mathbb{P}[h > Ce^{\alpha r^\beta}] && (6.7)\\ &= 1 - \int_{r>0} \mathbb{E}_I[\exp(-Ce^{\alpha r^\beta})|r, I] f_r(r)\mathrm{d}r && (6.8)\\ &= 1 - \int_0^\infty \mathcal{L}_I(Ce^{\alpha r^\beta}) f_r(r)\mathrm{d}r, && (6.9) \end{aligned}$$

where $f_r(r) = 2\pi r\lambda_s e^{-\pi\lambda_s r^2}$ is the pdf of the distance between the typical user and nearest cell, $h \sim \exp(1)$, and $\mathcal{L}_I(s)$ is the Laplace transform of the ICI,

$$\mathcal{L}_{\mathcal{I}}(s) = \exp\left(-\frac{2\pi\lambda_a}{\beta\alpha^{\frac{2}{\beta}}} \int_0^{e^{-\alpha r^\beta}} \frac{s}{1+sx} (-\ln(x))^{\frac{2-\beta}{\beta}} \mathrm{d}x\right), \tag{6.10}$$

where $\lambda_a = (1 - f_N(0))\lambda_s$ is the density of active cells. $\mathcal{L}_I(s)$ can be further simplified for the special case when $\beta = \frac{2}{n+1}$ such that

$$\mathcal{L}_I\{Ce^{\alpha r^\beta}\} = \exp\left(\sum_{k=0}^{n} \lambda_a r^{\frac{2k}{n+1}} \delta_k(C)\right), \tag{6.11}$$

$$\delta_k(C) = \frac{\pi(n+1)!}{k!\alpha^{n-k+1}} \frac{-C}{\Gamma(n-k+1)} \int_0^\infty \frac{x^{n-k}}{e^x + C}\mathrm{d}x. \tag{6.12}$$

Substituting (6.10) or (6.11) (in case of $\beta = \frac{2}{n+1}$) into (6.9), $F_c(C)$ is obtained. The probability $p = F_c(T)$ and $1-p$ that a randomly selected user belongs to either *Group2* or *Group1*, respectively. We define the number of users per cell in *Group1* as N_1 and in *Group2* as N_2.

Lemma 6.1 *For spatially distributed SCs and devices as two independent HPPPs with an average density of users per cell λ and assuming that the users are divided into two groups based on a normalized SIR threshold T, then, the CDF of the number of NOMA pairs per cell (Z) is given by:*

$$F_Z(z) = F_{N_1}(z) + F_{N_2}(z) - F_{N_1,N_2}(z,z), \tag{6.13}$$

$$F_{N_1}(n_1) = \sum_{i=0}^{n_1} \sum_{n=i}^{\infty} \binom{n}{i} (1-p)^i p^{n-i} f_N(n), \tag{6.14}$$

$$F_{N_2}(n_2) = \sum_{i=0}^{n_2} \sum_{n=i}^{\infty} \binom{n}{i} p^i (1-p)^{n-i} \quad f_N(n), \tag{6.15}$$

$$F_{N_1,N_2}(n_1,n_2) = \sum_{i=0}^{n_1} \sum_{j=0}^{n_2} \binom{i+j}{i,j} (1-p)^i p^j \ f_N(i+j), \tag{6.16}$$

where F_{N_1}, F_{N_2} are the CDFs of number of users in Group1 and Group2, respectively, and F_{N_1,N_2} is their joint CDF.

Proof: Given that the total number of users per cell N is known to be n, one can model N_1 and N_2 as binomially distributed with parameters $(n, 1-p)$ and (n, p), respectively. Hence, the conditional pmf of N_1 is given as:

$$f_{N_1|N}(n_1|n) = \binom{n}{n_1} (1-p)^{n_1} (p)^{n-n_1} \tag{6.17}$$

and (6.14) follows directly by summing over all possible values of n. Similarly, (6.15) can

be obtained. Taking into consideration that N_1 and N_2 are dependent, and given that the total number of users per cell N is known to be n, we can model their conditional joint pmf as a multinomial distribution with parameters (n, p_1, p_2) such that:

$$f_{N_1,N_2|N}(n_1, n_2|n) = \binom{n}{n_1, n_2} p_1^{n_1}(p_2)^{n_2}(1-p_1-p_2)^{n-n_1-n_2}, \tag{6.18}$$

where $\binom{n}{n_1,n_2} = \frac{n!}{n_1!n_2!(n-n_1-n_2)}$ is the multinomial coefficient. Thus, the joint CDF when the two groups are complementary (adjacent) can be expressed as in (6.16). ■

Theorem 6.1 *The network rate gain per cell γ of the hybrid NOMA/OMA scheme over the pure OMA scheme for an average density of devices per cell λ is given by*

$$\gamma = \frac{2}{\lambda}(\gamma_1 + \gamma_2) \times \mathbb{E}[Z], \tag{6.19}$$

$$\gamma_1 = \log_2\left(\frac{1+\frac{\varepsilon T}{1+\zeta(1-\varepsilon)T}}{\sqrt{1+T}}\right) + \int_{\log_2(1+\frac{\varepsilon T}{1+\zeta(1-\varepsilon)T})}^{\log_2(1+\frac{\varepsilon}{\zeta(1-\varepsilon)})} \frac{1-F_c\left(\frac{2^t-1}{\varepsilon-\zeta(1-\varepsilon)(2^t-1)}\right)}{1-p}dt - \frac{1}{2}\int_{\log_2(1+T)}^{\infty} \frac{1-F_c(2^t-1)}{1-p}dt, \tag{6.20}$$

$$\gamma_2 = \int_0^{\log_2(\frac{1+T}{1+\varepsilon T})} 1 - \frac{F_c\left(\frac{2^t-1}{(1-\varepsilon)-\varepsilon(2^t-1)}\right)}{p}dt - \frac{1}{2}\int_0^{\log_2(1+T)} 1 - \frac{F_c(2^t-1)}{p}dt, \tag{6.21}$$

where γ_1 and γ_2 are the rate gains of NOMA over OMA in Group1 and Group2, respectively. Z is the number of NOMA pairs per cell whose CDF is given in (6.13) and $\mathbb{E}[Z] = \sum_{allz}(1 - F_Z(z))$.

Proof: In the hybrid NOMA/OMA scheme, *Group1* and *Group2* are not necessarily of the same size. Hence, some devices will not be paired and will adopt OMA instead of NOMA. For a fair comparison, only half of the bandwidth allocated for a NOMA cluster will be allocated for those OMA devices and the rate gain of NOMA over OMA for the

users in *Group1* can be found as:

$$\gamma_1 = \mathbb{E}\left[\log_2(1+\frac{\varepsilon c_1}{\zeta(1-\varepsilon)c_1+1}) - \frac{1}{2}\log_2(1+c_1)\right]. \tag{6.22}$$

The first term in the expectation can be calculated as follows:

$$\begin{aligned}
&\mathbb{E}\left[\log_2(1+\frac{\varepsilon c_1}{\zeta(1-\varepsilon)c_1+1})\right] = \int_{t>0}\mathbb{P}\left[\log_2(1+\frac{\varepsilon c_1}{\zeta(1-\varepsilon)c_1+1}) > t\right]\mathrm{d}t \\
&= \int_{t>0}\mathbb{P}\left[c_1 > \frac{2^t-1}{\varepsilon-\zeta(1-\varepsilon)(2^t-1)}\right]\mathrm{d}t \\
&= \int_0^{\log_2(1+\frac{\varepsilon}{\zeta(1-\varepsilon)})} 1-F_{c_1}\left(\frac{2^t-1}{\varepsilon-\zeta(1-\varepsilon)(2^t-1)}\right)\mathrm{d}t \\
&\overset{(a)}{=} \log_2\left(1+\frac{\varepsilon T}{1+\zeta(1-\varepsilon)T}\right) + \int_{\log_2(1+\frac{\varepsilon T}{1+\zeta(1-\varepsilon)T})}^{\log_2(1+\frac{\varepsilon}{\zeta(1-\varepsilon)})} \frac{1-F_c\left(\frac{2^t-1}{\varepsilon-\zeta(1-\varepsilon)(2^t-1)}\right)}{1-p}\mathrm{d}t,
\end{aligned} \tag{6.23}$$

where $F_{c_1}(C) = \frac{F_c(C)-F_c(T)}{1-p}$ if $C > T$ and $F_{c_1}(C) = 0$ otherwise, is the CDF of the normalized SIR of the users in *Group1* and (a) follows directly from this. The second term in the expectation can be obtained directly by setting $\varepsilon = 1$. Similarly, for the devices in *Group2*,

$$\gamma_2 = \mathbb{E}\left[\log_2(1+\frac{(1-\varepsilon)c_2}{\varepsilon c_2+1}) - \frac{1}{2}\log_2(1+c_2)\right], \tag{6.24}$$

$$\begin{aligned}
\mathbb{E}\left[\log_2(1+\frac{(1-\varepsilon)c_2}{\varepsilon c_2+1})\right] &= \int_0^{\log_2(\frac{1}{\varepsilon})} 1-F_{c_2}\left(\frac{2^t-1}{(1-\varepsilon)-\varepsilon(2^t-1)}\right)\mathrm{d}t \\
&\overset{(a)}{=} \int_0^{\log_2(\frac{1+T}{1+\varepsilon T})} 1-\frac{F_c\left(\frac{2^t-1}{(1-\varepsilon)-\varepsilon(2^t-1)}\right)}{p}\mathrm{d}t
\end{aligned} \tag{6.25}$$

where (a) follows from the CDF of the normalized SIR of users in *Group2*, $F_{c_2}(C) = \frac{F_c(C)}{p}$ when $C < T$ and $F_{c_2}(C) = 1$ otherwise. The second term in the expectation in (6.24) can be obtained by setting $\varepsilon = 0$. As a result, to evaluate the network rate gain, the achievable gain of NOMA over OMA must be multiplied by the ratio of users adopting

NOMA relative to the total number of users, namely, $2\mathbb{E}(\frac{Z}{N})$. To proceed with the analysis, we approximate $\mathbb{E}[\frac{Z}{N}] \approx \frac{\mathbb{E}[Z]}{\mathbb{E}[N]}$ and $\mathbb{E}(N) = \lambda$ for a large enough number of cells and realizations. ■

To investigate the effect of the UDN environment, we steer our focus to ASE as a performance metric given in (6.3). Hence, the ASE gain obtained from the hybrid NOMA/OMA scheme compared to OMA is given in the following corollary.

Corollary 6.1 *The ASE gain of the hybrid NOMA/OMA scheme over pure OMA is given by:*

$$\gamma_{ASE} = p_a \times \lambda_s \times \gamma \tag{6.26}$$

where p_a is the cell activation probability, λ_s is the density of SCs, and γ is rate gain per cell given in (6.19).

6.4 Simulation Results

The actual cell rate gain given in (6.19) is reported using both simulation and analysis versus different system parameters. Unless otherwise stated, the SEPL parameters are set to $\alpha = 0.94$ and $\beta = 1/2$. The obtained results show perfect matching between the Monte-Carlo simulations and the derived analytical expressions. It is noteworthy that the actual rate gain calculated via simulations is obtained via averaging over the rates of all the users existing in the network either adopting NOMA or OMA. This proves that the approximation of $\mathbb{E}[\frac{Z}{N}] \approx \frac{\mathbb{E}[Z]}{\mathbb{E}[N]}$ is acceptable. Also, in the obtained results we adopt a perfect SIC scenario.

Fig. 6.2 shows the number of NOMA users per cell versus the SIR threshold T for different values of SC densities (λ_s). The maximum number of users adopting NOMA is expected when *Group1* and *Group2* have statistically equal sizes. Hence, we note from Fig. 6.2 that the value of the SIR threshold T which maximizes the number of NOMA

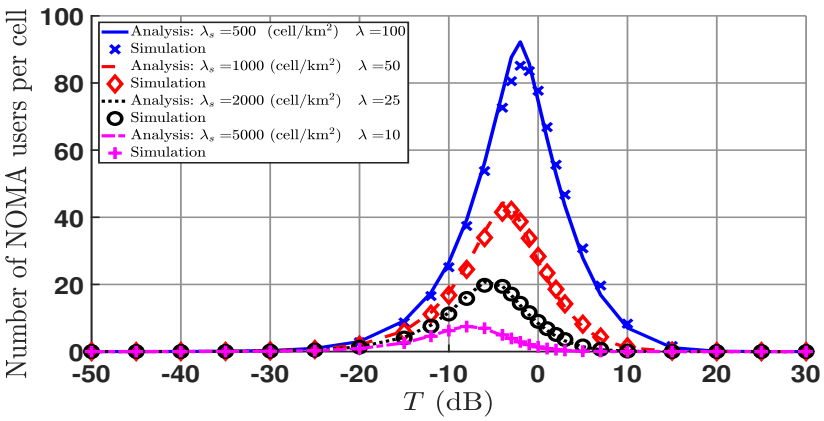


Figure 6.2: The average number of NOMA user per cell $2 \times Z$ versus the SIR threshold T, for different values of Small Cell densities (λ_s), MTCD density of $\lambda_m = 0.5$ devices/m^2, and activation probability of $\rho_m = 0.1$.

devices per cell yields $F_c(T) = 0.5$ for the corresponding system parameters. However, the motivation behind optimizing the value of T is to maximize the achievable rate per cell and ASE, not the number of NOMA devices.

Hence, Fig. 6.3 shows the achievable ASE of the hybrid NOMA/OMA scheme for different values of power allocation factors ε and SC densities (λ_s). Clearly, the performance of the hybrid NOMA/OMA scheme is a function of the threshold T which we need to optimize. In the remaining results reported in Figs. 6.4 and 6.5, we report the performance of the hybrid NOMA/OMA scheme at the optimum value of the threshold T obtained via numerical calculations.

From the obtained results in Fig. 6.3, we can see that increasing the density of SCs to double its value increases the ASE significantly for either the hybrid NOMA/OMA scheme or the pure OMA scheme. However, this increase is not linear with the SC density (λ_s). Although the hybrid NOMA/OMA scheme achieves a significant ASE gain compared to pure OMA, this gain diminishes with either increasing SC densities (λ_s) or decreasing the power allocation factor ε. Given this observation, one can see that NOMA becomes no longer valuable when the SC density increases in terms of the achievable ASE. Such behavior is clearly illustrated in Fig. 6.5 where the ASE gain diminishes with the SC

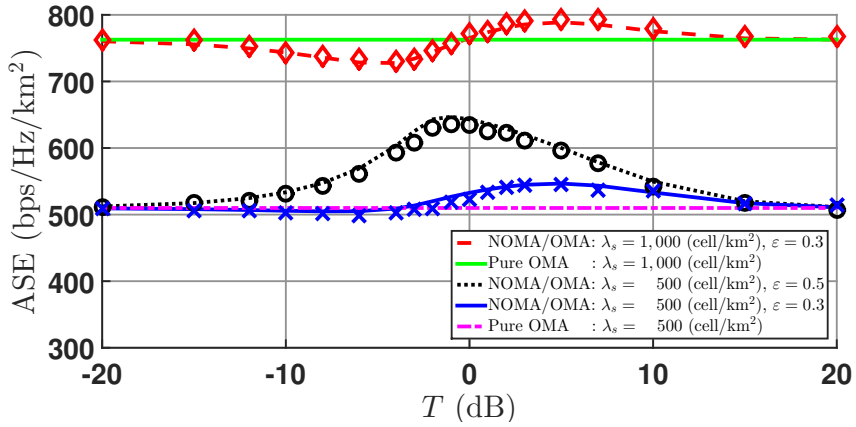


Figure 6.3: The network ASE of the hybrid NOMA/OMA scheme and pure OMA scheme versus the SIR threshold T, for different λ_s and ε. (Lines correspond to analysis and markers correspond to simulation.)

density (λ_s). However, NOMA can still be of great significance if the number of users to be simultaneously served exceeds the number of available resources per cell.

In Fig. 6.4, we show that the achievable ASE through both hybrid NOMA/OMA (with optimized SIR threshold T) and pure OMA schemes increases with the SC density (λ_s), which motivates the concept the UDN. The SEPL parameters α and β which describe the nature of the practical environment significantly affect the performance of the network which becomes obvious with higher λ_s. It is noteworthy that a real scenario with actual measurements reported in [21] is approximately represented by the settings of $\alpha = 0.94$ and $\beta = 1/2$.

In Fig. 6.6, we optimize the SIR threshold T to maximize the ASE gain of the network. At this optimal T, we report the percentage gains of the users in *Group1* and *Group2* separately, considering only users adopting NOMA in each group. We notice that within the range of $0.2 < \varepsilon < 0.45$, the users in both groups achieve gain when deploying NOMA.

It is noteworthy that if we adopt the standard path loss model, the achievable rates per cell for either NOMA or OMA schemes will not vary with the density of SCs. This happens due to the fact that the gain obtained in the useful signal power due to shorter

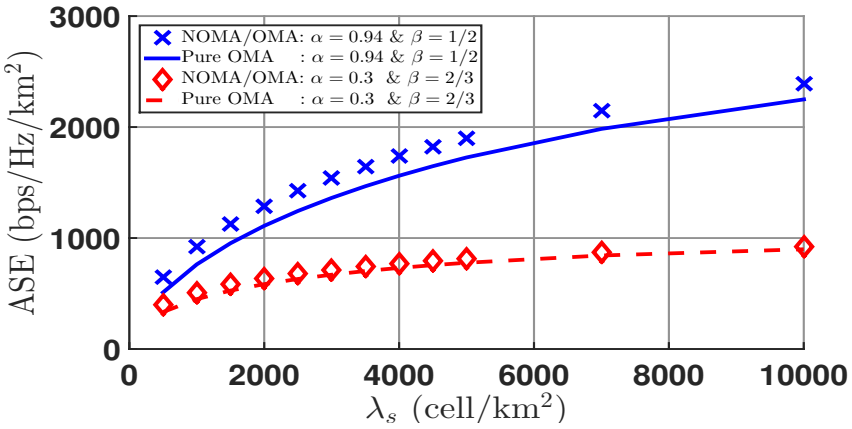


Figure 6.4: Achievable ASE of the optimized NOMA/OMA and pure OMA schemes versus SC density (λ_s), for different environmental parameters of the SEPL model and power allocation factor $\varepsilon = 0.5$.

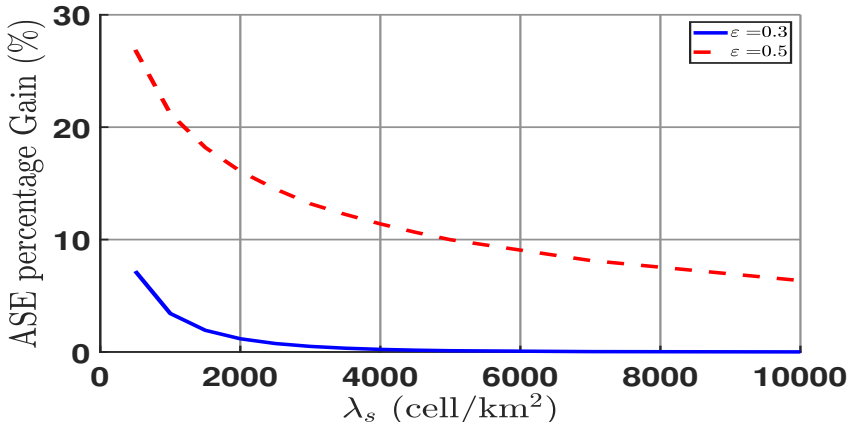


Figure 6.5: The percentage ASE gain of the optimized NOMA/OMA over pure OMA versus λ_s, for different power allocation factors (ε).

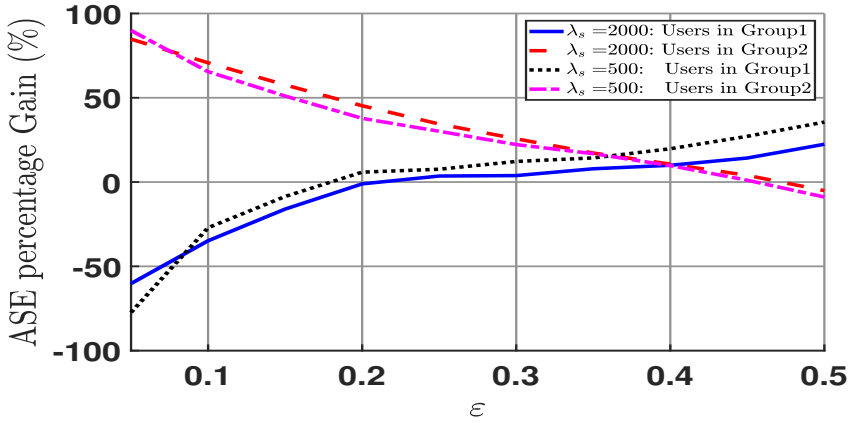


Figure 6.6: The percentage ASE gain for users in *Group1* and *Group2* of the optimized NOMA/OMA versus ε, for different Small Cell densities (λ_s).

distances will be exactly eliminated by the increase of the ICI. Consequently, the ASE increases linearly with the density of SCs (λ_s) in the case of a path loss model. However, for a practical UDN scenario, the Stretched Exponential Path Loss yields the achievable rates per cell to decrease as the ICI dominates the power gain in the useful signal, hence, the ASE increases sub-linearly with the SC density.

6.5 Summary

A hybrid NOMA/OMA scheme for multiple access along with a suggested clustering scheme based on SIR is proposed. We provided analytical expressions for the network ASE gain over pure OMA. Our results show that an optimum value of the SIR threshold exists such that the gain from the hybrid NOMA/OMA scheme is maximized where this optimum value is a function of power allocation ratios, SEPL parameters, and SC density as well as MTCDs density. Finally, we show that the gain obtained from NOMA vanishes with the increasing density of SCs and that a range of the NOMA power allocation ratio exists such that the performance of all users is lower bounded by the achievable performance in the case of OMA.

Chapter 7

Supporting mMTC in mmWave-UDNs with Wireless Backhauling

7.1 Introduction

In the same scope as the previous chapter, we tackle mMTC use case targeted in 5GB networks. In doing so, we exploit two leading technologies in 5G, namely, UDN and mmWave. On one hand, in a UDN, a large number of SCs exists in the network benefiting the MTCDs by reducing the number of MTCDs associated with the same SC and by providing a shorter serving distance to the associated SC. However, the sensed ICI dramatically increases. On the other hand, mmWave can help mitigate severe ICI thanks to the high penetration losses experienced in this frequency band. Besides, the gigantic available bandwidths in the mmWave band provide each SC with a higher number of Resource Blocks (RBs) that can be utilized to support more MTCDs per SC. Taking into consideration the challenge of supporting the backhaul of the massive number of SCs, we consider a massive Multiple-Input Multiple-Output (mMIMO) microwave backhauling

system between the SC and the Central Processing Units (CPUs) of the core network that may be attached to a second tier of macro Base Stations (BSs). Using tools from stochastic geometry, we provide analytical expressions for the achievable performance in terms of density of supported MTCDs, achievable SINR per MTCD, and capacity per SC under the considered limited wireless backhaul scenario. The main contributions can be summarized as follows:

- To fully exploit the available cellular network resources in 5G and beyond, we jointly consider UDN and mmWave to support MTC. In particular, both severe ICI of UDNs and high penetration losses due to blockages in mmWave band are considered.
- Considering a practical use case, we study the performance under a limited wireless backhaul scenario while taking into account the correlated effects of different system parameters on both AL and BH.
- We propose a sub-optimal calculation of SC density, at which the achievable gains from the deployed SCs in terms of the limited capacity are maximized. Accordingly, we offer network operators a reasonable estimation of the density of SCs that should be activated.
- We formulate a novel mathematical framework and derive analytical expressions for the density of supported MTCDs, the achievable SINR per MTCD, and the average capacity per SC (for both AL and BH) using tools from stochastic geometry. Besides, we derive a tight lower bound of the achievable bounded capacity.

7.1.1 Related Works

MTC supported by UDN has been widely considered in the literature as an inevitable use case targeted in 5G and beyond [10, 12, 19, 23, 64, 105]. In [11], the authors provided an extensive survey on the opportunities and challenges facing MTC while supported by Non-Orthogonal Multiple-Access (NOMA) and UDN. For instance, the authors

in [105] considered a downlink scenario for an ultra-dense user-centric IoT network supported by non-ideal optical fiber backhaul links. The authors aimed at minimizing energy consumption over the whole network and investigated the significance of the non-ideal backhaul on achievable performance. In the same scope, the authors in [23] considered coexistent HTCUs and MTCDs in a UDN under limited backhaul capacity for the HTC. it is noteworthy that the authors therein did not specify the type of the backhauling system.

Considering unlimited backhaul capacity, the authors in [12] considered a scenario of coexistent HTCUs and MTCDs within a UDN and sharing the available spectrum based on a NOMA scheme. The authors investigated the downlink capacity performance of both types of users/devices while adopting a partial idle-mode activation scenario of certain SCs. In particular, two association schemes were considered for MTCDs, namely, connect-to-active (C2A) and connect-to-closest (C2C) where MTCDs either associate with the closest already-activated SC by an HTCU or associate with the closest SC in general. A trade-off between shorter serving distance and increased ICI was investigated therein. In [10], however, the authors adopted a hybrid NOMA/OMA approach to support MTCDs in an UDN. The works in [19,64] considered the uplink performance of MTC either solely or coexistent with HTC with unlimited backhaul capacity.

In the above-mentioned works, the utilized spectrum is the traditional microwave band. Advancing into the mmWave band, many factors have to be reconsidered, ahead of which is the blockage effect on signal propagation. In this regard, many works in the literature adopted tools from stochastic geometry to formulate a mathematical framework for the achievable performance in the mmWave band. For instance, in [106], the authors considered a realistic measurement-based mmWave channel models [107] and proposed a Signal-to-Interference plus Noise Ratio (SINR) coverage analysis framework based on different antenna radiation patterns. By doing so, the authors defined the aligned gain corresponding to the desired signal and the misaligned gain corresponding to the interfering signals. The obtained results therein showed that, with reasonable accuracy, both

aligned and misaligned gains combined with the channel gain can be approximated by exponential distributions such that their corresponding means depend solely on the number of transmitting and receiving antennas. Hence, we adopt this approximation in our considered system model.

In a different scope, the work in [108] considered a mmWave network and investigated the blockage effect on the served devices. In particular, the authors proposed equipping some blockages with Reconfigurable Intelligent Surfaces (RISs) to achieve higher coverage probability and reduce the blind-spot areas of devices, *i.e.,* the areas at which the devices are in outage due to lack of LoS SCs. Similar to [106], the authors adopted tools from stochastic geometry to derive an analytical framework for the achievable performance. The work in [109] studied the uplink performance of cell-free massive MIMO under a mmWave fronthaul network. It is assumed that the network includes multiple CPUs to which APs are associated in a distance-based criterion. Also, each AP communicates with its serving CPU using point-to-point communication, however, it neglects the impact of interference on the system performance on the access link between the AP and user equipment. Tools from stochastic geometry are applied to derive closed-from expressions for the average achievable uplink rates. In general, interested readers may refer to [110, 111] for more detailed analyses of blockages based on a stochastic geometry approach.

The rest of the chapter is organized as follows. Section 7.2 describes the system model. In Section 7.3, we perform the required mathematical analysis for the AL and the BH. Section 7.4 includes the simulation setup and shows the obtained results along with their discussions and insights.

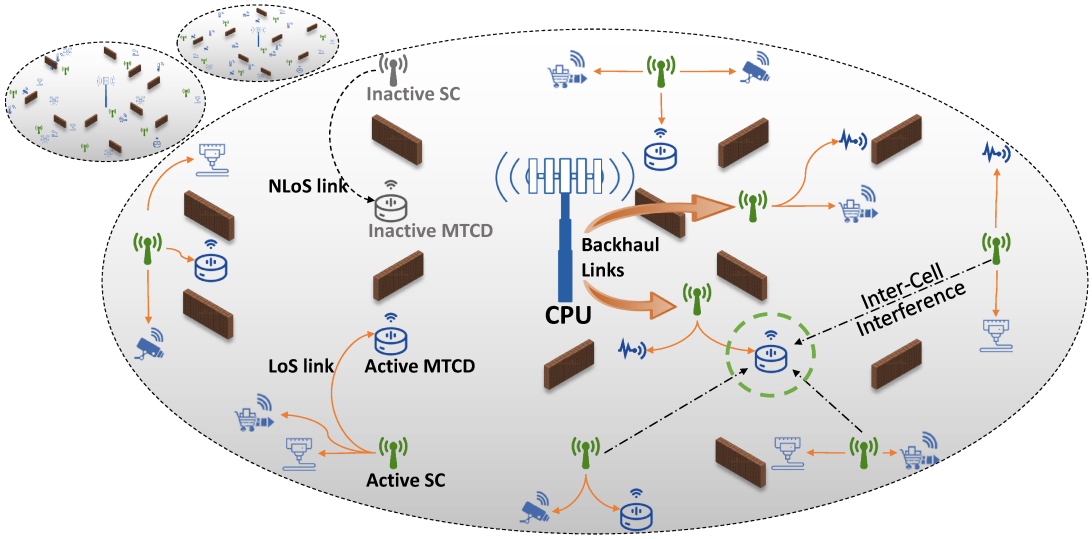


Figure 7.1: A realization of mmWave UDN supporting MTC under limited wireless backhaul capacity and dense blockages.

7.2 System Model

We consider the downlink traffic of MTC in an UDN environment with dense blockages such as an indoor environment. Besides, we consider a typical use case of 5G which is mMTC, in which a massive number of MTCDs are to be served [1]. In particular, each MTCD tries to associate with the SC providing the strongest average received power as illustrated in Fig.7.1. Given that the number of MTCDs is relatively large compared to the number of SCs, we can not make use of the idle mode capability utilized in UDNs [4], in which, SCs with no associated users/devices are switched off. Hence, to mitigate the effect of ICI and to fully exploit the available network resources, we use the mmWave band for the communications between the SCs and the massive number of MTCDs to be served, denoted as the Access Link (AL). By doing so, we benefit from the excessive available bandwidth, the high directional beamforming, and the high penetration loss experienced in the mmWave band [112]. In particular, we positively make use of the existing blockages in the considered indoor environment to reduce the amount of severe ICI found in UDNs.

In our model, we assume that both SCs and MTCDs are uniformly distributed according to two independent HPPPs Φ_s and Φ_m with corresponding densities λ_s and λ_m,

respectively. For a practical scenario, we assume an activity profile for the MTCDs with activation probability ρ_m such that only a fraction ρ_m of the MTCDs are simultaneously active [10]. Hence, we assume a thinned HPPP $\Phi_m^a \subset \Phi_m$ with density $\lambda_m^a = \rho_m \lambda_m$ to represent the set of active MTCDs at certain time instant. It is noteworthy that typical densities of SCs and active MTCDs would result $\lambda_m^a \gg \lambda_s$ and yields almost all SCs to be active.

To support the massive number of MTCDs, we assume that the available frequency/time resources are divided into a number N_{RB} of orthogonal RBs. Then, each SC can serve at most N_{RB} MTCDs within a certain TTI across the whole available bandwidth for the access link. In particular, for a given RB, we assume that an SC can serve at most one MTCD. Hence, each SC causes ICI towards its neighboring SCs on each of its utilized RBs. To further mitigate ICI and to enhance the received signal power, we assume that each SC is equipped with a number of transmitting antennas N_s which are used to deploy directional beamforming towards the served MTCD. The authors in [106] used the measurement-based mmWave channel model of the New York University (NYU) Wireless Group [107] and applied some curve-fitting techniques to approximate the combined channel power gain and antenna gain $|h^{AL}|^2$ as follows for both aligned and misaligned scenarios:

$$|h_q^{AL}|^2 \sim \exp(1/\mathcal{A}_q), \quad q \in \{alg, mis\}, \tag{7.1}$$

where $\{alg, mis\}$ refer to the aligned and misaligned scenarios, respectively. The average combined channel and antenna gain $\mathcal{A}_q$ is given in [106] by

$$\mathcal{A}_{alg} = N_s, \tag{7.2}$$

$$\mathcal{A}_{mis} = 1/\sin^2\left(\frac{3\pi}{2\sqrt{N_s}}\right), \tag{7.3}$$

where N_s is the number of transmitting antennas while assuming that each antenna el-

ement adopts an isotropic radiation pattern. In the same scope, the beamwidth of the directed beam is $\theta_{alg} = 2\pi/\sqrt{N_s}$ which corresponds to the aligned scenario while the misaligned beamwidth becomes $\theta_{mis} = 2\pi - \theta_{alg}$.

As mentioned earlier, the mmWave signals suffer from severe penetration loss when passing through physical blockages. In our model, we assume that the centers of blockages are distributed according to another independent HPPP Φ_b with density λ_b while the blockage lengths are assumed to be distributed uniformly between L_{min} and L_{max} with an average blockage length of $\bar{L} = (L_{min}+L_{max})/2$ while the widths of blockages are neglected [108]. According to [110], one can approximate the blockage effect by an exponentially decaying function such that the probability of a LoS link between any two nodes separated by a distance d is

$$P_{LoS}(d) = \exp(-\mathcal{B}_b d), \tag{7.4}$$

where $\mathcal{B}_b = \frac{2\lambda_b \bar{L}}{\pi}$. Hence, the probability of non-LoS link between the two nodes is $P_{NLoS}(d) = 1 - P_{LoS}(d)$. It is noteworthy that the above assumption of exponential blockage does not take into account the correlation effect between two nodes that are in close proximity to each other. However, such an effect can be neglected [108], especially when the density of SCs is relatively large as in the considered UDN environment.

For the experienced large-scale fading in the AL, the state of the link $z \in \{LoS,\ NLoS\}$ yields different path loss intercepts β_z^{AL} and path loss exponents α_z^{AL} depending on the operating frequency band. Based on actual propagation measurements for mmWave communications [107, 113], the path loss between a typical transmitter-receiver pair is

$$\zeta_z^{AL} = \beta_z^{AL}\ d^{-\alpha_z^{AL}}, \quad z \in \{LoS, NLoS\}, \tag{7.5}$$

where d is the distance between the two nodes and β_z^{AL} is the path-loss at a 1 m distance. Although it is a common assumption in the literature that a UDN environment is interference-limited, we can not follow such an assumption herein due to the blockage

effect in the mmWave band. In other words, the ICI may reduce to very small values compared to thermal noise under certain circumstances such as high density of blockages compared to the density of SCs. Hence, throughout our investigation, both interference and noise are considered in both simulation and analysis.

Moving to the backhaul (BH) links between the SCs and their serving CPUs, we consider a wireless mMIMO-based backhaul network in which each SC associates to its closest CPU over a wireless backhaul link. As we consider an indoor deployment of UDN system, we assume that there are no LoS links between SCs and their serving CPUs which constitutes a worst-case scenario for the backhaul communications. In that, the channel vector between the SC s and its serving CPU $\mathrm{c_s}$ is modeled as a Rayleigh fading channel as follows:

$$\boldsymbol{g}_{c_s s}^{BH} = \sqrt{\zeta_{c_s s}^{BH}} \boldsymbol{h}_{c_s s}^{BH}, \tag{7.6}$$

where $\boldsymbol{h}_{c_s s}^{BH}$ represents the backhaul small-scale fading channels vector whose entries are i.i.d $CN(0,1)$. Also, $\zeta_{c_s s}^{BH}$ reflects the backhaul large-scale fading channel coefficient which can be calculated as its counterpart in the access link, but with a different reference path-loss β_z^{BH} and different path-loss exponent α_z^{BH}. Different from the access link communication, and due to the high losses accompanying signal transmission over mmWave band under NLoS conditions, we assume that the communication links between SCs and CPUs are carried out over the microwave band. In this regard, we consider the microwave band around $f_{carrier}^{BH} = 6.5\ GHz$ is adopted for the backhaul communication due to the high available bandwidth in this frequency band that can handle the data transfer in the backhaul network [114]. Co-located mMIMO-based data transmission is considered for the communication between each CPU and its associated SCs. To this end, we assume that each CPU is equipped with N_c microwave antennas. Also, each SC is equipped with a dedicated microwave antenna for the backhaul communication with its serving CPU. Taking advantage of the multiple antennas at CPU, and considering superior number of CPU antennas to the associated number of SCs, we adopt zero-forcing (ZF) beam-

forming/detection technique for data transmission/reception between each CPU and its associated SCs [115]. More specifically, ZF precoding is applied at CPUs to transmit the data of its associated SC over the backhaul network. Also, ZF detection is applied at each CPU to detect the transmitted pilots from its associated SCs over the backhaul network.

7.3 Performance Analysis

In this section, we derive analytical expressions for the achievable SINR and the average achievable ASE under the considered bounded wireless backhaul scenario. We begin by assuming a typical MTCD located at the origin. This typical MTCD, according to Slivnyak's theorem, reflects the average achievable performance of all MTCDs when the SCs are distributed according to a PPP. The typical MTCD tries to associate with the SC that provides the strongest average received power. In this case, the tagged SC can either have a LoS or NLoS link with the typical MTCD. It is noteworthy that such SC would not necessarily be the closest one in terms of distance. However, given the high penetration loss experienced in the mmWave band, we assume that power received from a SC with a LoS will always dominate. Given that the CDF of the distance between the typical MTCD and its closest LoS SC is

$$F_r^{LoS}(r) = 1 - \exp\left(-\frac{2\pi\lambda_s}{\mathcal{B}_b^2}(1-(1+\mathcal{B}_b r)e^{-\mathcal{B}_b r}\right), \tag{7.7}$$

hence, the pdf of the distance between the typical MTCD and its tagged LoS SC, given that the LoS exists, can be expressed as

$$f_{r|LoS}^{LoS}(r) = \frac{2\pi\lambda_s r}{1-p_m^{out}} \exp\left(-\mathcal{B}_b r - 2\pi\lambda_s \frac{1-(1+\mathcal{B}_b r)e^{-\mathcal{B}_b r}}{\mathcal{B}_b^2}\right), \tag{7.8}$$

where p_m^{out} is the LoS SC outage probability for the typical MTCD calculated as

$$p_m^{out} = F_r^{LoS}(\infty) = \exp(\frac{-2\pi\lambda_s}{\mathcal{B}_b^2}). \tag{7.9}$$

Then, the probability of the typical MTCD being associated with an SC would be

$$p_m^{assoc} = 1 - p_m^{out}. \tag{7.10}$$

7.3.1 Density of Served MTCDs

Similar to (7.9), the probability that a SC does not have any associated MTCD can be expressed as $p_s^{out} = \exp(\frac{-2\pi\lambda_m^a}{\mathcal{B}_b^2})$. It is clear that such probability approaches zero when either the density of active MTCDs λ_m^a is very high or the blockage parameter $\mathcal{B}_b$ is relatively small. In addition, in case of no blockage, the idle mode probability of a SC defined as the probability of no associated MTCDs is $p_s^{idle} = (3.5\lambda_s/(3.5\lambda_s + \lambda_m^a))^{3.5}$. Although, the idle mode probability of SCs plays a vital role in an UDN environment when calculating the ICI, in our system model, such probability can be neglected due to the very high density of active MTCDs compared to the density of SCs. Hence, the density of active SCs can be approximated by $\lambda_s^a \approx (1 - p_s^{out})\lambda_s$.

Then, to find the density of supported MTCDs, we need to calculate the number of supported MTCDs by each active SC. First, we find the distribution of the number of MTCDs N associated to the SC serving the typical MTCD. In doing so, we refer to [10], in which the required pmf of N is given by

$$f_N(n) = \frac{(\frac{35}{\lambda+35})^{3.5}}{\Gamma(3.5)} \left(\frac{\lambda}{\lambda+3.5}\right)^n \frac{\Gamma(n+3.5)}{\Gamma(n+1)}; \quad n \geq 0, \tag{7.11}$$

where $\lambda = p_m^{assoc}\lambda_m^a/\lambda_s^a$ is the average number of active and non-outage MTCDs associated to an active SC and $\Gamma(.)$ is the gamma function defined as $\Gamma(x) = \int_0^\infty t^{x-1}e^{-x}\mathrm{d}x$. Then, the

number of supported MTCDs M_m per an active SC would be given by $M_m = \min(N, N_{RB})$. Since, $w = \min(x, y)$ yields $F_w(w) = F_x(w) + F_y(w) - F_{x,y}(w, w)$, then, the distribution of M_m, given that N and N_{RB} are independent and N_{RB} is constant, is

$$F_{M_m}(m) = \begin{cases} F_N(m) = \sum_{n=0}^{m} f_n(n) & ; \ 0 \leq m < N_{RB} \\ 1 & ; \ m \geq N_{RB} \end{cases}, \tag{7.12}$$

$$f_{M_m}(m) = \begin{cases} f_N(m) & ; \ 0 \leq m < N_{RB} \\ \sum_{n=N_{RB}}^{\infty} f_n(n) & ; \ m = N_{RB} \\ 0 & ; \ m > N_{RB} \end{cases}. \tag{7.13}$$

Accordingly, the density of served MTCDs would be obtained as

$$\lambda_m^{supp} = \lambda_s^a \times \mathbb{E}[M_m] = \lambda_s^a \times \sum_{m=0}^{N_{RB}} m f_{M_m}(m), \tag{7.14}$$

where $\bar{M}_m = \mathbb{E}[M_m]$ is the average number of supported MTCDs per SC. Hence, the probability that an active MTCD is allocated a RB by its associated SC is

$$p_m^{alloc} = \frac{\bar{M}_m}{N_{RB}}. \tag{7.15}$$

7.3.2 Achievable SINR on Access Link

In this subsection, we provide the main mathematical analysis for the AL. An active MTCD is denoted as a supported MTCD when it associated to a SC (with probability p_m^{assoc} given in (7.10)) and allocated a RB (with probability p_m^{alloc} given in (7.15)). Afterwards, the achievable SINR by the typical supported MTCD located at the origin would be

$$\gamma_0^{AL} = \frac{S_0}{I_0 + \ \sigma^2 \times W_{RB}^{AL}}, \tag{7.16}$$

where S_0 is the useful signal power received at the typical MTCD, $I_0 = I_0^{alg} + I_0^{mis}$ is the sensed ICI generated from other active LoS SCs which are either aligned or misaligned towards the typical MTCD, σ^2 is the power density of the thermal noise at the receiver, and $W_{RB}^{AL} = W_{AL}/N_{RB}$ is the available AL bandwidth per RB. Following the previous discussions, the useful signal power at the typical MTCD can be represented as

$$S_0 = P_s^0 \ \beta_{LoS}^{AL} \ |h_{alg}^{AL}|^2 \ r_0^{-\alpha_{LoS}^{AL}}, \tag{7.17}$$

where P_s^0 is the transmission power level of the serving SC of the typical MTCD, β_{LoS}^{AL} and α_{LoS}^{AL} are the path loss intercept and path loss exponent given in (7.5), respectively, $|h_{alg}^{AL}|^2$ is combined aligned channel power and antenna gain given in (7.1), and r_0 is the distance between the typical MTCD and its serving SC for which its distribution is given in (7.8).

Lemma 7.1 *The CCDF of the received signal power at the typical supported MTCD can be obtained as follows*

$$\mathbb{P}[S_0 > S_{th}] = \int_0^{\infty} \exp\left(-\frac{S_{th} \ r^{\alpha_{LoS}^{AL}}}{P_s^0 \beta_{LoS}^{AL} \ \mathcal{A}_{alg}}\right) f_{r|LoS}^{LoS}(r) \ \mathrm{d}r. \tag{7.18}$$

Proof: Starting from (7.17), we have

$$\begin{aligned}\mathbb{P}[S_0 > S_{th}] &= \mathbb{E}_{\Phi_s^{LoS}}\left[\mathbb{P}\left(|h_{alg}^{AL}|^2 > \frac{S_{th} \ r_0^{\alpha_{LoS}^{AL}}}{P_s^0 \ \beta_{LoS}^{AL}}\right) \ \middle| \ r_0\right], \\ &\overset{a}{=} \mathbb{E}_{\Phi_s^{LoS}}\left[\exp\left(-\frac{S_{th} \ r_0^{\alpha_{LoS}^{AL}}}{P_s^0 \ \beta_{LoS}^{AL} \ \mathcal{A}_{alg}}\right) \ \middle| \ r_0\right],\end{aligned} \tag{7.19}$$

where Φ_s^{LoS} is the thinned PPP of LoS active SCs with density $\lambda_s^{LoS}(d) = P_{LoS}(d) \times \lambda_s^a$ and (a) is obtained from the exponential distribution of $|h_{alg}^{AL}|^2$ given in (7.1). Then, (7.18)

follows directly by using the pdf of the serving distance of the LoS SC given in (7.8). ■

Next, we move to deriving the CCDF of the achievable SINR for the typical supported MTCD. Following (7.16), we provide the SINR coverage probability in the following theorem.

Theorem 7.1 *The SINR coverage probability of supported MTCDs, i.e., MTCDs which are associated with a LoS SC and allocated a RB, is given by*

$$\mathbb{P}[\gamma^{AL} > \gamma_{th}] = \int_0^{\infty} e^{-\frac{\gamma_{th}\sigma^2 W_{RB}^{AL} r^{\alpha_{LoS}^{AL}}}{P_s^0 \ \beta_{LoS}^{AL} \ \mathcal{A}_{alg}}} \prod_{q\in\{alg,mis\}} \mathcal{L}_{I_q}\left(\frac{\gamma_{th}\ r^{\alpha_{LoS}^{AL}}}{P_s^0 \beta_{LoS}^{AL} \mathcal{A}_{alg}}\right) \quad f_{r|LoS}^{LoS}(r)\ \mathrm{d}r, \tag{7.20}$$

$$\mathcal{L}_{I_q}\left(\frac{\gamma_{th}\ r^{\alpha_{LoS}^{AL}}}{P_s^0 \beta_{LoS}^{AL} \mathcal{A}_{alg}}\right) = \exp\left(-\pi\ p_m^{alloc}\ p_q \lambda_s^a\ (\frac{\gamma_{th}\mathcal{A}_q}{\mathcal{A}_{alg}})^{\frac{2}{\alpha}}\ r^2 \int_{(\frac{\gamma_{th}\mathcal{A}_q}{\mathcal{A}_{alg}})^{-2/\alpha}}^{\infty} \frac{e^{-\mathcal{B}_b(\frac{\gamma_{th}\mathcal{A}_q}{\mathcal{A}_{alg}})^{1/\alpha}\ r\ \sqrt{u}}}{1+u^{\frac{\alpha}{2}}}\mathrm{d}u\right), \tag{7.21}$$

where $q \in \{alg, mis\}$ and $p_{alg} = \theta_{alg}/2\pi,\ p_{mis} = 1 - p_{alg}$ are the probabilities of the interferer being aligned or misaligned, respectively. p_m^{alloc} is the probability that the considered RB is being utilized given in (7.15).

It is noteworthy that for the inner integral given in (7.21) to be numerically solvable, we had to approximate the variable (r) by stair function with fixed value through each step.

Proof: Appendix B.3 ■

Having the SINR distribution at hand, we move to the next step which is deriving the distribution of the AL achievable capacity per SC. First, we define the achievable rate by the typical MTCD as [116],

$$\tau(\gamma') = \begin{cases} \tau' & ;\gamma^{AL} \geq \gamma' \\ 0 & ;otherwise \end{cases}, \tag{7.22}$$

where $\tau' = W_{RB}^{AL} \log_2(1+\gamma')$ and γ^{AL} is the achievable SINR. $\mathbb{P}[\gamma^{AL} > \gamma']$ can be obtained from (7.20) where γ' is determined based on the receiver sensitivity and the required performance in terms of achievable rate and bit error rate. Hence, the achievable rate per typical MTCD $\tau(\gamma')$ is a random variable that follows a Bernoulli distribution with parameter $p = \mathbb{P}[\gamma^{AL} > \gamma']$, however, it is multiplied by a constant value of $\tau' = W_{RB}^{AL} \log_2(1+\gamma')$. Accordingly, for a fixed number of supported devices per SC, let it be ${M_m}'$, the number of MTCDs achieving non-zero rate would follow a binomial distribution with parameters p and ${M_m}'$. However, the number of supported devices per SC M_m, whose distribution is given in (7.13), is not fixed. From the definition of a Bernoulli distribution, the CCDF of $\tau(\gamma')$ is

$$\mathbb{P}[\tau(\gamma') > \tau_{th}] = \begin{cases} 1 & ; \tau_{th} < 0 \\ \mathbb{P}[\gamma^{AL} > \gamma'] & ; 0 \leq \tau_{th} < \tau' \\ 0 & ; \tau_{th} \geq \tau' \end{cases} . \tag{7.23}$$

Hence, the distribution of the achievable capacity on the AL per SC is given in the following theorem.

Theorem 7.2 *The CCDF of the achievable capacity per active SC over the access link is*

$$\mathbb{P}[C_s^{AL} > C_{th}] = \sum_{m=1}^{N_{RB}} \mathbb{P}\left[\tau(\gamma') > \frac{C_{th}}{m}\right] \mathbb{P}[M_m = m], \tag{7.24}$$

where the CCDF of $\tau(\gamma')$ is given in (7.23) and the distribution of M_m is given in (7.13).

Proof: The achievable capacity per SC on the AL is determined by the number of supported devices per SC M_m and ow many of them satisfy the threshold SINR value γ'. Since the achievable rate by each of the supported devices is $\tau(\gamma')$, then, the total achievable capacity of a specific SC will be $C_s^{AL} = \tau(\gamma') \times M_m$. Hence, the CCDF of C_s^{AL}

is

$$\mathbb{P}[C_s^{AL} > C_{th}] = \mathbb{E}_{M_m}\Big[\mathbb{P}\big[\tau(\gamma') \times M_m > C_{th} \mid M_m\big]\Big],$$
$$= \mathbb{E}_{M_m}\Big[\mathbb{P}\big[\tau(\gamma') > \frac{C_{th}}{M_m} \mid M_m\big]\Big], \tag{7.25}$$

which directly leads to (7.24) by summing over all possible values of $M_m \geq 1$ and multiplying by its pmf correspondence given in (7.13), hence, completes the proof. Note that for $M_m = 0$, this means no associated MTCDs, *i.e.,* $N = 0$, and the SC would be inactive.

■

Consequently, one can obtain the average achievable capacity per active SC over the access link C_s^{AL} as follows,

$$\bar{C}_s^{AL} = \int_0^{\infty} \mathbb{P}[C_s^{AL} > c]\mathrm{d}c,$$
$$\overset{(a)}{=} \sum_{m=1}^{N_{RB}} \mathbb{P}[M_m = m] \int_0^{m\tau'} \mathbb{P}[\gamma^{AL} > \gamma']\mathrm{d}c,$$
$$= p\tau' \sum_{m=1}^{N_{RB}} m\mathbb{P}[M_m = m], \tag{7.26}$$

where $p = \mathbb{P}[\gamma^{AL} > \gamma']$, $\tau' = W_{RB}^{AL}\ \log_2(1+\gamma')$, and (a) is obtained from (7.23).

7.3.3 Backhaul Communication

During the backhaul communication, a certain CPU should convey the corresponding symbols of served users by one of its associated SCs. Consequently, each CPU is required to send different symbols to its associated SCs over the backhaul network. This problem can be considered as a beamforming problem in which each CPU exploits the multiple antennas at its side to direct the different data to different SCs simultaneously over the same time-frequency resources. In this regard, we assume that each CPU has perfect channel state information of its associated SCs. This can be justified as both SCs and CPUs are deployed at fixed locations with high-altitude antenna elements of the CPU.

This leads to wireless backhaul links with long channel coherence time in which a perfect channel estimation of the deployed SCs can be readily obtained. Moreover, to assure a good backhaul performance, we apply ZF precoding for data transmission between CPU and its associated SCs. Let us consider $\jmath_s$ as the composite signal to be transmitted to the s^{th} SC from its serving CPU. Thus, the received signal at SC s is given by

$$\breve{\jmath}_s = \sqrt{P_c} \sum_{s' \in \mathcal{M}_c} \sqrt{\breve{\gamma}_{s'}} (\boldsymbol{g}_{c_s s}^{BH})^T \boldsymbol{\mathcal{Q}}_{s'} \jmath_{s'} + w_s, \tag{7.27}$$

where P_c denotes the DL transmission power at CPUs and $\mathcal{M}_c$ being the set of associated SCs to the CPU c. In addition, w_s reflects the additive thermal noise at the s^{th} SC. The precoding vector $\boldsymbol{\mathcal{Q}}_{s'}$ for the data to be transmitted to SC $s' \in \mathcal{M}_c$ is given by

$$\boldsymbol{\mathcal{Q}}_{s'} = \left(\boldsymbol{G}_{c_s}^{BH}\right)^* \Big(\left(\boldsymbol{G}_{c_s}^{BH}\right)^T \left(\boldsymbol{G}_{c_s}^{BH}\right)^* \Big) \mathsf{Inv} \boldsymbol{e}_{s'}, \tag{7.28}$$

where $\boldsymbol{G}_{c_s}^{BH}$ represents the backhaul channel matrix between the CPU c_s and its associated SCs. Also, $\boldsymbol{e}_{m'}$ denotes the m'^{th} column in the identity matrix $\boldsymbol{I}_{M_{c_s}}$ where M_{c_s} represents the number of associated SCs to the CPU c_s. The term $\breve{\gamma}_{s'}$ denotes the power normalization factor for the precoding vector of SC $s' \in \mathcal{M}_{c_s}$ that can be calculated as follows:

$$\breve{\gamma}_{s'} = \frac{1}{M_{c_s} \mathbb{E}\Big\{ \Big[\Big(\left(\boldsymbol{G}_{c_s}^{BH}\right)^T \left(\boldsymbol{G}_{c_s}^{BH}\right)^* \Big) \mathsf{Inv} \Big]_{(s',s')} \Big\}}. \tag{7.29}$$

Thus, the received signal at the s^{th} SC can be rewritten as follows:

$$\breve{\jmath}_s = \sqrt{P_c \gamma_s} \jmath_s + w_s. \tag{7.30}$$

Capitalizing on Lemma 2.9 in [117] for $M_c \times M_c$ central complex Wishart matrix with N_c degree of freedom which satisfies $N_c \geq M_c + 1$, the power normalization factor for the

precoding vector of SC s' can be calculated as follows:

$$\breve{\gamma}_{s'} = \frac{\left(N_c - M_{c_s}\right)\zeta_{c_s s}^{BH}}{M_{c_s}}. \tag{7.31}$$

Consequently, the provided backhaul capacity in bps/Hz to SC s from its serving CPU will be

$$C_s^{BH} = W^{BH}\log_2\left(1 + \frac{P_c\left(N_c - M_{c_s}\right)\zeta_{c_s s}^{BH}}{M_{c_s} W^{BH}(\sigma_w^J)^2}\right), \tag{7.32}$$

where N_c is the number of transmitting antennas per CPU, M_{c_s} is the number of associated SC to the CPU, and W^{BH} is available bandwidth for backhaul communications.

Theorem 7.3 *The CCDF of the achievable capacity per active SC over the backhaul link is*

$$\mathbb{P}[C_s^{BH} > C_{th}] =$$

$$\sum_{m_c=1,2,\ldots}^{N_c-1}\left(1 - \exp\left(-\pi\lambda_c\left(\frac{P_c\beta^{BH}(N_c - m_c)}{m_c W^{BH}(\sigma_w^J)^2\left(2^{C_{th}/W^{BH}} - 1\right)}\right)^{\frac{2}{\alpha^{BH}}}\right)\right)\mathbb{P}\left[M_{C_s} = m_c\right], \tag{7.33}$$

where the pmf of the number of SCs associated to the CPU M_{C_s} can be obtained from (7.11) by setting $\lambda = \lambda_s^a/\lambda_c$.

Proof: Appendix B.3 ■

Similar to (7.26) but adopting a different approach for the continuous random variable C_s^{BH}, the average capacity per SC over the backhaul link can be obtained as

$$\bar{C}_s^{BH} = \int_0^\infty \mathbb{P}[C_s^{BH} > c]\mathrm{d}c,$$

$$= \sum_{m_c=1,2,\ldots}^{N_c-1}\mathbb{P}\left[M_{C_s} = m_c\right]\int_0^\infty\left(1 - \exp\left(-\pi\lambda_c\left(\frac{P_c\beta^{BH}(N_c - m_c)}{m_c W^{BH}(\sigma_w^J)^2\left(2^{c/W^{BH}} - 1\right)}\right)^{\frac{2}{\alpha^{BH}}}\right)\right)\mathrm{d}c. \tag{7.34}$$

Finally, the achievable capacity per active SC is the minimum capacity achievable on both access link and backhaul link, *i.e.,*

$$C_s^{bounded} = \min\left(C_s^{AL}, C_s^{BH}\right), \tag{7.35}$$

$$\mathbb{P}[C_s^{bounded} > C_{th}] = \mathbb{P}[C_s^{AL} > C_{th}, C_s^{BH} > C_{th}]. \tag{7.36}$$

It is noteworthy that C_s^{AL} and C_s^{BH} are not independent as they both depend on the density of active SCs. However, for tractability, we will assume independence between them to achieve an approximation for the achievable capacity per SC. In particular,

$$\begin{aligned}
\bar{C}_s^{bounded} &= \int_0^{\infty} \mathbb{P}[C_s^{bounded} > c]\mathrm{d}c, \\
&\approx \sum_{m_c=1,2,\ldots}^{N_c-1} \sum_{m=1,2,\ldots}^{N_{RB}} \mathbb{P}\left[M_{C_s} = m_c\right] \mathbb{P}[M_m = m] \\
&\times \int_0^{\infty} \left(1 - \exp\left(-\pi\lambda_c \left(\frac{P_c \beta^{BH}(N_c - m_c)}{m_c W^{BH} (\sigma_w^J)^2 \left(2^{c/W^{BH}} - 1\right)}\right)^{\frac{2}{\alpha^{BH}}}\right)\right) \mathbb{P}\left[\tau(\gamma') > \frac{c}{m}\right] \mathrm{d}c, \\
&= \sum_{m_c=1,2,\ldots}^{N_c-1} \sum_{m=1,2,\ldots}^{N_{RB}} \mathbb{P}\left[M_{C_s} = m_c\right] \mathbb{P}[M_m = m] \\
&\times \; p \; \times \int_0^{m\tau'} \left(1 - \exp\left(-\pi\lambda_c \left(\frac{P_c \beta^{BH}(N_c - m_c)}{m_c W^{BH} (\sigma_w^J)^2 \left(2^{c/W^{BH}} - 1\right)}\right)^{\frac{2}{\alpha^{BH}}}\right)\right) \mathrm{d}c,
\end{aligned} \tag{7.37}$$

where last step is obtained similar to (7.26) such that $p = \mathbb{P}[\gamma^{AL} > \gamma']$ and $\tau' = W_{RB}^{AL} \log_2(1 + \gamma')$.

7.4 Simulation Results

In order to verify the accuracy of the derived analytical expressions, we perform Monte-Carlo simulations and compare the results from both simulations and analysis. In the shown figures, lines indicate analysis and markers represent simulations unless

mentioned otherwise. Thousands of spatial realizations are performed and averaged over an area of 1×1 km^2. The simulation parameters are chosen to reflect practical scenarios. Unless otherwise stated, the density of MTCDs is set as one device per square meter. Amongst the existing MTCDs, only a fraction $\rho_m = 10\%$ are assumed to be active and having data ready to be received. The number of transmitting antennas at the CPU and SC are $N_c = 250$ and $N_s = 10$, respectively. The operating center frequency of the AL is $f_c^{AL} = 73$ GHz and the available bandwidth is $W^{AL} = 1$ GHz with number of RBs $N_{RB} = 10$ or 100. For the BH, the operating frequency is $f_c^{BH} = 6.5$ GHz with allocated bandwidth of 200 MHz. The noise power is set to -174 dBm/Hz. The density of blockages is assumed $1,000$ blockages per km^2 with uniform length distribution between 5 m and 15 m. The SINR threshold is set to $\gamma' = 0$ dB. The path loss exponents are chosen as $\alpha_{LoS}^{AL} = 3.1$ and $\alpha^{BH} = 3$ to reflect an urban environment. The corresponding path loss intercepts are $\beta_{LoS}^{AL} = 20\log_{10}(4\pi d_0 f_c^{AL}/(c))$ and $\beta^{BH} = 32.4 + 20\log_{10}(f_c^{BH} \times 10^{-9})$ where $c = 3 \times 10^8$ is the speed of light [113].

7.4.1 Density of supported MTCDs

In Fig.7.2, we show the effect of deploying more SCs on the density of MTCDs that have LoS links with at least one SC. An MTCD with no LoS SC is considered at outage. It can be seen that within an environment with dense blockages, increasing the density of SCs to around 0.3 of the density of blockages, given the considered distribution of blockage lengths, leads to a scenario where almost all MTCDs have an associated LoS SC, *i.e.,* no outage. However, depending on the number of RBs available at each SCs, only limited number of such MTCDs with associated SCs can be supported.

Accordingly, in Fig.7.3, we show the density of MTCDs with allocated RBs. For those MTCDs without allocated RBs at certain time interval, they can be prioritized in the next time interval to be allocated a RB, however, such investigation of scheduling problem is beyond the scope of this work. In Fig.7.3, we investigate the impact of increasing SC

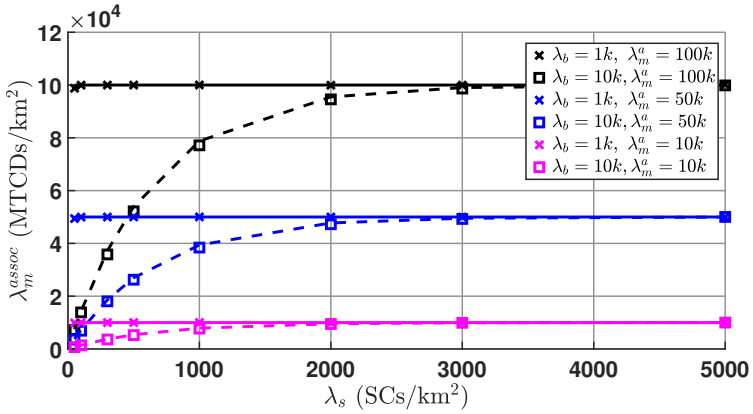


Figure 7.2: Density of MTCDs associated to a LoS SC.

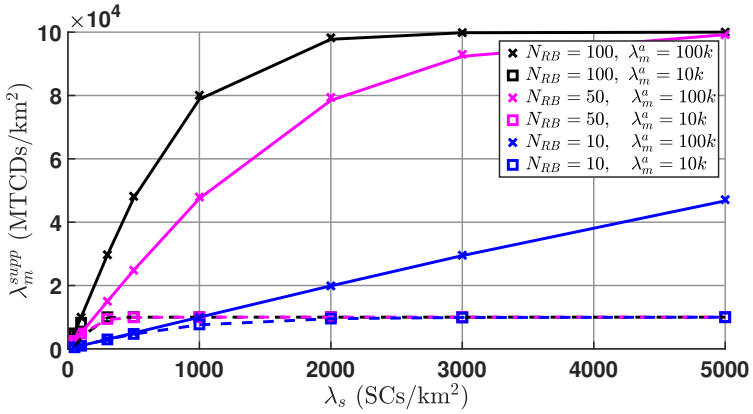


Figure 7.3: Density of MTCDs with an allocated RB.

density under different combinations of active MTCD density and available number of RBs at each SC. One can expect and notice the positive impacts of increasing both the density of SCs (λ_s) as well as the number of RBs (N_{RB}) available at each SC. However, it should also be noted that increasing both λ_s and N_{RB} may become useless beyond certain limits depending on the density of active MTCDs λ_m^a. For example, in both top and bottom curves of Fig.7.3, having a total density of RBs equal to the density of active MTCDs (at $\lambda_s = 1,000$ SCs/km^2), *i.e.*, $N_{RB} \times \lambda_s = \lambda_m^a$, would yield only 80 percent of the active MTCDs to be allocated RBs.

On the other hand, increasing such total density of RBs to triple the density of active MTCDs leads to all MTCDs being allocated RBs. Interestingly, one can also notice that the impact of increasing the number of RBs per SC is equivalent to increasing the SC

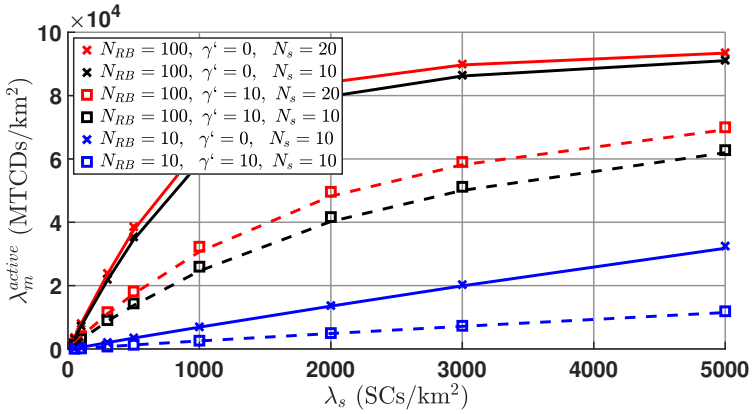


Figure 7.4: Density of SINR-active MTCDs, achieving SINR above the predetermined threshold (γ'). Markers indicate simulation and lines represent analysis.

density on the density of MTCDs with allocated RBs. However, such trade-off between density of SCs and number of RBs per SC would lead to different results in terms of the achievable SINR as can be seen in next figures.

In Fig.7.4, we show the impacts of three different system parameters, namely, the number of transmitting antennas per SC used to deploy beamforming in the mmWave band N_s, the number of RBs available at each SC N_{RB}, and the predetermined SINR threshold used to define the modulation scheme and transmission rate for each MTCD γ'. First, similar to λ_m^{assoc} in Fig.7.2 and λ_m^{alloc} in Fig.7.3, increasing the both λ_s and $N_R B$ leads to higher density of SINR-active MTCDs (λ_m^{active}), however, until all MTCDs are SINR-active. Interestingly, it can be noted that that when the product $N_{RB} \times \lambda_s$ is fixed, the case with larger N_{RB} yields a slightly higher density of SINR-active MTCDs. This can be resolved to the lesser ICI sensed when the number of RBs is large which yields more MTCDs achieving the threshold SINR over their links. However, the allocated bandwidth for each RB is smaller which means lower achievable rate by each MTCD. Second, increasing N_s should improve received signal powers and reduce the ICI sensed at the MTCDs, which can be easily noted under all scenarios. Finally, although increasing the value of γ' would yield higher achievable rate per active MTCD as given in (7.22), it yields significantly lower density of SINR-active MTCDs λ_m^{active}.

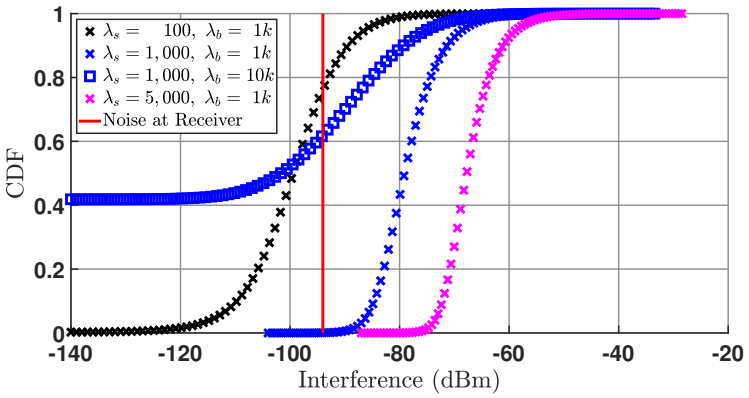


Figure 7.5: CDF of the sensed ICI obtained through Mont-Carlo simulations and compared to the thermal noise level to show interference-limited regime versus noise-limited regime ($N_{RB} = 10$).

7.4.2 Achievable SINR on the Access Link

In a different scope, we show in Fig.7.5 how does the system move from noise-limited regime to interference-limited regime and vice versa under different system conditions. The thermal noise at the receiver is fixed and depends on the reception bandwidth which in turn depends on the available bandwidth and number of orthogonal RBs. In Fig.7.5, we can see that for a relatively high SC density λ_s and moderate density of blockages, the system is interference-limited. However, when either the density of SC is relatively small and/or the density of blockages (or equivalently, their average lengths) is high, the system becomes noise-limited. Hence, such behavior confirms the importance of considering both interference and noise when investigating the the performance of our system model. Since there is no analytical expression for the ICI, the different curves shown therein are obtained only through Mont-Carlo simulations.

Consequently, Fig.7.6 shows the accuracy of the obtained analytical expression of the achievable SINR over the mmWave AL given in Theorem 7.1 under different combinations of the AL system parameters. It can be noted that increasing λ_s leads to a slight improvement in the SINR thanks to the reduced serving distance and higher received power from the associated SC compared to and despite the higher sensed ICI. Also,

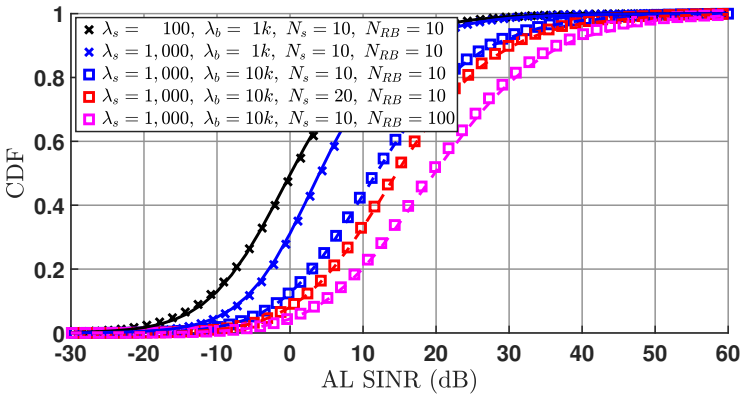


Figure 7.6: CDF of the achievable SINR of active MTCDs with allocated RBs. Markers indicate simulation and lines represent analysis.

increasing the blockage density λ_b leads as well to a higher SINR due to the reduction in the sensed ICI. Intuitively, increasing the number of transmitting antennas N_s helps to improve the achievable SINR. Finally, increasing the number of RBs available at each SC N_{RB} yields significant SINR improvement for two reasons. First, the thermal noise level decreases with the smaller bandwidth for each RB. Second, with higher N_{RB} and the fact that active MTCDs are randomly allocated the RBs, less number of RBs become utilized which significantly reduces the amount of sensed ICI.

7.4.3 Average Capacity per SC

In Fig.7.7, we investigate the effect of different system parameters on the average capacity per SC between the SC itself and its served MTCDs, referred to as the Access Link (AL). First, consider the black curve with x-markers ($\gamma' = 0$, $N_{RB} = 100$, and $N_s = 10$) as a reference set of parameters. In particular, one can notice the existence of an optimal λ_s that maximizes the AL average capacity per SC. When λ_s is small, the serving distances of MTCDs towards their serving SCs are relatively large, almost all available RBs will be utilized, and the system lies in the noise-limited regime. While λ_s slightly increases, the system remains in the noise-limited regime as explained in Fig.7.5, however, the SCs start to get closer to the MTCDs which enhances the received signal power at MTCDs, the

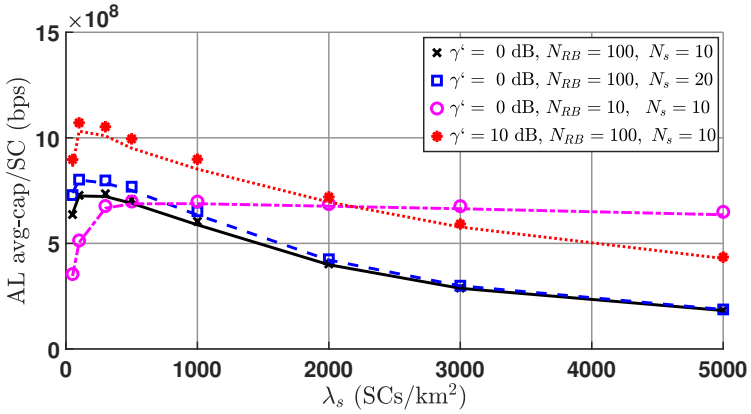


Figure 7.7: Average achievable capacity per SC on the Access Link (AL). Markers indicate simulation and lines represent analysis.

achievable SINR per MTCD, and the number of SINR-active MTCDs per SC. Hence, the AL average capacity per SC increases in return. As λ_s increases beyond certain limit, the system switches to the interference-limited regime and no more gain can be achieved in terms of the SINR since the gain in the received signal power would be mitigated by higher interference levels. Moreover, as λ_s further increases, the number of MTCDs associated with a SC decreases. When the available RBs per SC become unutilized, a reduction in the AL average capacity per SC is encountered.

Increasing the number of transmitting antennas per SC N_s would improve the AL average capacity per SC, however, with the added cost of the additional antennas to be deployed. It is important to notice for large λ_s, the achievable gain from increasing N_s gets negligible since more RBs become unutilized as shown by the square markers in Fig.7.7. In a different scope, increasing the value of the predetermined SINR threshold γ' leads to decreasing the number of SINR-active MTCDs per SC while increasing the achievable rate by each of those SINR-active MTCDs. The combined effects of the latter can be seen to increase the total capacity per SC. In other words, it can be deduced from the right-most curve in Fig.7.6 that the probability of MTCDs being SINR-active would reduce from around 0.95 to only 0.8 when γ' changes 0 dB to 10 dB. On the other hand, the average achievable rate per MTCD would increase to $\log_2(11)/\log_2(2)$ of its initial value as can

be derived from (7.22).

Finally, reducing the number of RBs per SC in Fig.7.7, *i.e.,* $N_{RB} = 10$ instead of $N_{RB} = 100$ results in some changes in the achievable performance. On one hand, when the system operates in the noise-limited regime (*i.e.,* λ_s is small), some degradation in the average capacity per SC can be noted due to the increased bandwidth per RB and consequently the higher value of thermal noise. On the other hand, when the system operates in the interference-limited regime (*i.e.,* λ_S is large), having fully utilized RBs, even for larger densities of SCs, yields higher capacity per SC. However, such a scenario means smaller density of supported MTCDs as shown in Fig.7.3 and Fig.7.4.

It is noteworthy that deploying more SCs is accompanied with additional costs. Hence, maximizing the achievable capacity per SC would constitute a tempting and non-negligible objective for network designers. Besides, it is to be noted that maximizing the AL average capacity per SC would not necessarily lead to actual maximization of the bounded capacity given the limited wireless backhaul capacity which depends on λ_s as well. Consequently, in Fig.7.8 we show the average capacity per SC on the backhaul links. The obtained results show the positive gains acquired by increasing both the number of transmitting antennas per CPU N_c as well as the density of CPUs λ_c when the density of SCs is fixed. It can be noted that increasing the number of CPUs has a more significant effect than increasing the number of antennas per CPU on the BH average achievable capacity per SC, however, with a higher cost.

Fig.7.9 concludes the results obtained in Fig.7.7 and Fig.7.8. In particular, we investigate the effect of multiple system parameters on the actual achievable capacity per SC which is bottle-necked by both the AL capacity and BH capacity. In our mathematical analysis in the previous section, we were able to derive an analytical lower bound for the average bounded capacity per SC. The achievable instantaneous capacity per SC is the minimum between the AL and the BH, however, we show the average capacity per SC in Fig.7.9. In this regard, one can notice how tight is the derived lower bound from the

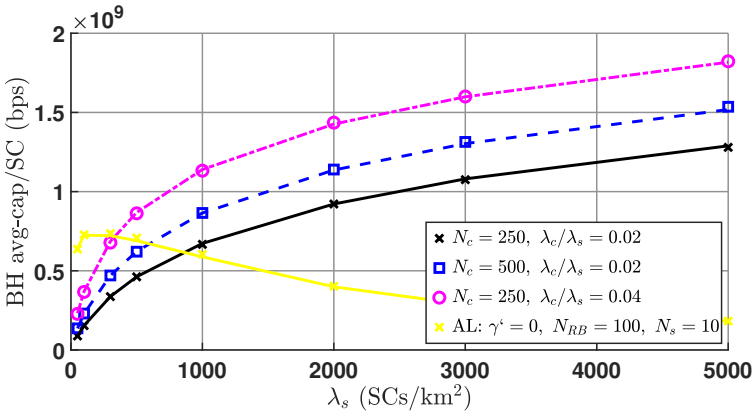
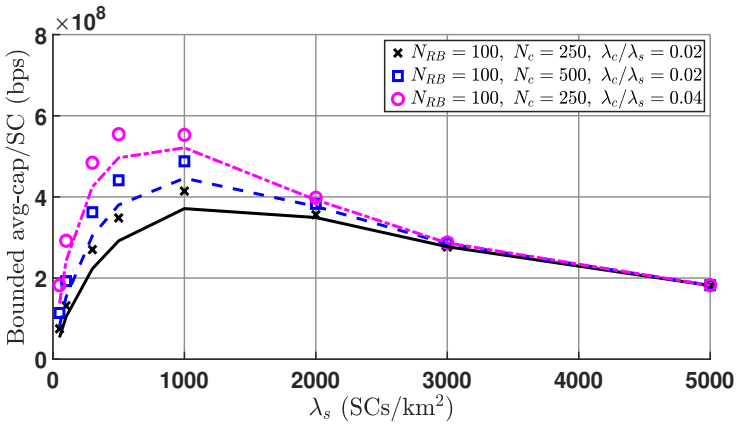


Figure 7.8: Average achievable capacity per SC on the Backhaul link (BH). Markers indicate simulation and lines represent analysis.

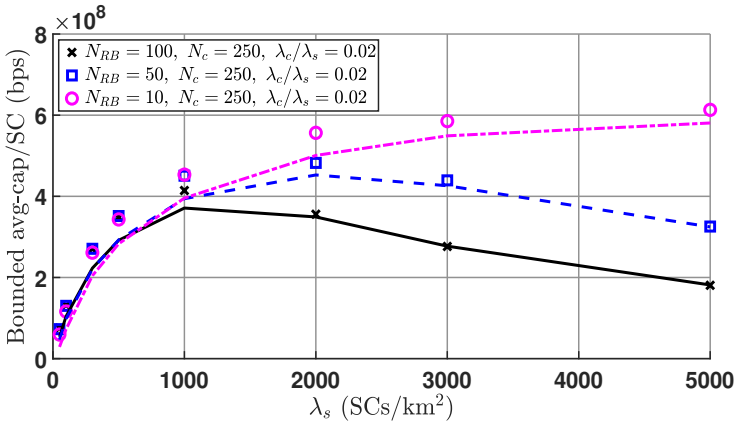
actual results obtained through heavy simulations. Besides, we show the existence of an optimal λ_s at which the network operator can achieve the maximum benefit from the deployed SCs.

7.5 Summary

In this chapter, we investigate the performance of massive Machine-Type Communication (mMTC) in an Ultra-Dense Network (UDN) environment while utilizing the millimeter Wave band and considering a wireless backhaul support for the SCs. In particular, we take into account the significant effect of blockages over mmWave signal propagation as well as the severe Inter-Cell Interference (ICI) of UDNs. In this regard, we show how the system may vary between interference-limited and noise-limited regimes based on the relative densities of small cells and blockages. Moreover, the implications of the limited capacities of the wireless backhaul links and the correlation between the access link and the backhaul link are investigated. The derived analytical expressions for the density of supported MTCDs and the average capacity per SC can be used to tune different system parameters to achieve enhanced performance. For instance, we show that an optimal value of SC density can be determined to maximize the utilization of the deployed SCs.



(a)



(b)

Figure 7.9: Average achievable bounded capacity per SC. Markers indicate simulation and lines represent the analytical lower bound.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

Ultra-Dense Network (UDN) is one of the leading technologies in 5G and beyond networks aiming at tackling different use cases such as enhanced mobile broadband (eMBB), ultra-reliable and low-latency communication (URLLC), and massive machine-type communication (mMTC). In this thesis, we explored some of the rising challenges in future cellular networks and proposed promising solutions within the UDN architecture while integrating other supportive technologies. In particular, we focused on supporting both human-type communication (HTC) and machine-type communication (MTC) either separately or simultaneously. For HTC, we targeted providing high data rates, supporting computation-intensive applications, and securing data transfer between the users and their associated small cells (SCs). For MTC, we aimed at supporting a massive number of devices simultaneously. In doing so, we took into account the accompanying challenges of UDNs such as the limitations imposed on supporting an enormous number of SCs with ideal backhaul links, the severe inter-cell interference (ICI) in UDNs, and the complexity and time associated with performing system-level simulations. Significant contributions have been made in this thesis, which is evident from the track record of publications that

resulted from this research work. In the following, we briefly conclude the work presented in this Ph.D. thesis:

1. The UDN architecture is used in 5G and beyond networks to support various use cases, including enhanced mobile broadband, ultra-reliable and low-latency communication, and massive machine-type communication.

2. Our research focused on supporting both human-type communication and machine-type communication separately or simultaneously.

3. To achieve high data rates for human-type communication, a multiple-association scheme was adopted where each user associates with multiple small cells in its vicinity. This scheme was shown to provide gains under limited backhaul capacity constraints, which are common in UDNs.

    - The *MultiCell* size, or the number of associated small cells, was shown to be able to be optimized for different system parameters such as backhaul capacities and densities of small cells and human-type users.

    - The potential gains that can be acquired by coexisting machine-type devices from the excess number of activated small cells were investigated. It was shown that a higher density of machine-type devices can be supported.

4. To reduce the total delay in processing computation-intensive tasks for human-type users, edge computing servers were integrated within the small cells of the UDN, and an offline adaptive task division technique was proposed.

    - The *MultiCell* size was found to need to be optimized in relation to different system parameters such as the densities of small cells and human-type users, and the processing capabilities of the human-type users and edge computing servers.

5. A multiple-association physical layer security protocol was proposed to increase the average secrecy rate per human-type user by splitting data traffic into multiple paths to provide spatial diversity against potential eavesdroppers.

    - The secrecy performance was shown to improve with increasing the *MultiCell* size and/or the density of small cells. However, higher densities of human-type users or eavesdroppers resulted in degradation of the average secrecy rate per user, which could be fully mitigated by increasing the *MultiCell* size.

6. A hybrid non-orthogonal multiple-access (NOMA)/orthogonal multiple-access (OMA) scheme was proposed for machine-type communication using a clustering process based on the signal-to-interference ratio (SIR).

    - An optimum value of the SIR threshold was found to exist that maximizes the gain from the proposed scheme, and this value was shown to be a function of the NOMA power allocation ratio, SEPL parameters, small cell density, and machine-type device density.
    - The gains obtained from NOMA were found to vanish with increasing density of small cells, where the performance of all users was lower bounded by the achievable performance in the case of OMA.

7. Our research also investigated a scenario of a UDN environment utilizing the millimeter wave band and considering a wireless backhaul of the small cells.

    - It was shown how the system can vary from an interference-limited regime to a noise-limited regime and vice versa based on the relative densities of small cells and blockages.
    - The implications of the limited capacities of the wireless backhaul links and the correlation between the access link and the backhaul link were also investigated.

- Different system parameters can be adjusted to achieve enhanced performance. For example, it was shown that an optimal value of small cell density can be determined to maximize the utilization of deployed small cells.

8.2 Future Work

In this section, we shed light on some directions of future research that can further extend the contributions of this thesis:

1. Stochastic geometry provided us with analytical expressions that can be used to tune several system parameters in order to enhance the achievable performance. However, resource allocation optimization in terms of frequency, power, and *MultiCell* size can help us to achieve further significant enhancements. It is noteworthy that the large scale of the number of nodes considered in this thesis requires careful selection of the optimization technique to provide a practical implementation. For example, machine learning techniques can be promising candidates.

2. Interference management is crucial for the deployment of UDN. ICI is seen as the dominant limiting factor of achievable performance due to the large number of SCs. In this thesis, we considered idle mode capability as one alternative for mitigating the ICI. However, with the massive number of MTCDs that may yield activating all SCs, other alternatives should be considered. The integration of a multiple-input multiple-output (MIMO) system within the UDN architecture can significantly boost the achievable performance and help to mitigate the severe ICI. A careful design of the MIMO system and resource allocation towards HTC and MTC should be investigated in future works.

3. In this thesis, we considered two use cases of 5G, namely, eMBB and mMTC with aid of UDNs and other integrated technologies. Future works may be directed towards

the third use case which is URLLC to show how the UDN resources can be ideally exploited to enhance the reliability and latency performance of certain users.

4. Integrated access and backhaul (IAB) in UDNs may be challenging due to the severe ICI in addition to the self-interference. However, it is impractical and/or infeasible to support the massive number of SCs in UDNs with optical fiber backhaul links. Further investigation should be allocated to IAB in UDNs with proper allocation of resources and careful handling of the accompanying interferences.

5. Energy consumption or energy efficiency is crucial for the environment and for limited-battery devices such as IoT devices. Special attention should be given to energy consumption in UDNs where the massive number of SCs can exaggerate the amount of energy dissipated into the network. On the other hand, with the SCs getting closer to the MTCDs, less power may be needed for uplink transmissions which helps to extend the battery lifetime of the MTCDs. The tradeoff between the energy consumption of SCs and that of MTCDs should be carefully investigated in future works.

Appendix A

List of Publication

List of Publications

Journal Articles

1. M. Elbayoumi, M. Kamel, W. Hamouda and A. Youssef, "NOMA-Assisted Machine-Type Communications in UDN: State-of-the-Art and Challenges," in *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 1276-1304, Secondquarter 2020.

2. M. Elbayoumi, W. Hamouda and A. Youssef, "Multiple-Association Supporting HTC/MTC in Limited-Backhaul Capacity Ultra-Dense Networks," in *IEEE Transactions on Communications*, vol. 69, no. 6, pp. 4113–4127, 2021.

3. M. Elbayoumi, W. Hamouda and A. Youssef, "Edge Computing and Multiple-Association in Ultra-Dense Networks: Performance Analysis," in *IEEE Transactions on Communications*, vol. 70, no. 8, pp. 5098-5112, 2022.

Conference Proceedings

1. M. Elbayoumi, M. Kamel, W. Hamouda and A. Youssef,"Capacity Analysis of Downlink NOMA-Based Coexistent HTC/MTC in UDN," *2019 International Conference on Communications, Signal Processing, and their Applications (ICCSPA)*, 2019, pp. 1-6.

2. M. Elbayoumi, W. Hamouda and A. Youssef,"A Hybrid NOMA/OMA Scheme for MTC in Ultra-Dense Networks," *2020 IEEE Global Communications Conference*, 2020, pp. 1-6.

3. M. Elbayoumi, W. Hamouda and A. Youssef, "Secrecy Performance in Ultra-Dense Networks with Multiple Associations," *2020 IEEE Globecom Workshops* , 2020, pp. 1-6.

4. M. Elbayoumi, W. Hamouda and A. Youssef, "Ergodic Secrecy Rate Analysis of Ultra-Dense Networks with Multiple Antennas," *2021 IEEE International Conference on Communications (ICC)*, 2021, pp. 1-6.

Appendix B

Proofs

B.1 Proofs of Chapter 3

Proof of Theorem 3.1

Given the equal bandwidth allocated to each tier, the instantaneous achievable rate by the typical user from its associated cell on *tier-k* given that it is not disconnected is given by

$$R_k^M = \frac{1}{M}\log_2(1+\gamma_k). \tag{B.1}$$

With γ_k given in (3.10), one can evaluate the Complementary Cumulative Distribution Function (CCDF) of γ_k as follows

$$\begin{aligned}
\mathbb{P}[\gamma_k > \zeta] &= \mathbb{E}_{I'_k, r_k}\left[\mathbb{P}\Big[h_k > \zeta(\frac{\sigma^2}{P} + I'_k)e^{\alpha r_k^\beta} \mid I'_k, r_k\Big]\right], \\
&\overset{(a)}{=} \mathbb{E}_{I'_k, r_k}\left[\exp\Big(-\zeta(\frac{\sigma^2}{P} + I'_k)e^{\alpha r_k^\beta}\Big)\right], \\
&= \int_{r>0} e^{-\zeta\frac{\sigma^2}{P}e^{\alpha r^\beta}}\, \mathbb{E}_{I'_k}\left[\exp\Big(-\zeta I'_k e^{\alpha r^\beta}\Big)|r\right] f_{r_k}(r)\, \mathrm{d}r, \\
&= \int_0^\infty e^{-\zeta\frac{\sigma^2}{P}e^{\alpha r^\beta}}\, \mathcal{L}_{I'_k}(\zeta e^{\alpha r^\beta}) f_{r_k}(r)\, \mathrm{d}r,
\end{aligned} \tag{B.2}$$

where $\mathbb{E}_X[.]$ denotes expectation over the random variable X and (a) follows from the unit mean exponential distribution of the channel gain h_k corresponding to the Rayleigh small scale fading. $\mathcal{L}_{I'_k}(s) = \mathbb{E}_{I'_k}[e^{-sI'_k}]$ is the Laplace transform of the normalized ICI on *tier-k*, I'_k, given in (3.11). Hence,

$$\begin{aligned}\mathcal{L}_{I'_k}(s) &= \mathbb{E}_{I'_k}\left[e^{-sI'_k}\right] \\ &= \mathbb{E}_{\Phi^a_k,h_j}\left[\exp\left(-s\sum_{j\in\Phi^a_k\setminus b_{k_0}} h_j e^{-\alpha r_j^\beta}\right)\right] \\ &= \mathbb{E}_{\Phi^a_k}\left[\prod_{j\in\Phi^a_k\setminus b_{k_0}} \mathbb{E}_h\Big[\exp(-s\ h\ e^{-\alpha r_j^\beta})\Big]\right]. \end{aligned} \tag{B.3}$$

where the last step is obtained given that all small scale fading coefficients are i.i.d.. From the probability generating function of an HPPP, one can find that [22]

$$\mathbb{E}\left[\prod_{z\in\Phi} f(z)\right] = \exp\left(-\lambda' \iint_{\mathbb{R}^2} (1-f(z))\mathrm{d}z\right), \tag{B.4}$$

where λ' is the density of the HPPP Φ and $\mathbb{R}^2$ is the two dimensional Euclidean space. Since the power of the channel gain, h, is exponentially distributed with mean μ, then,

$$\mathbb{E}_h\left[e^{-\varrho h}\right] = \frac{1}{1+\mu\varrho}. \tag{B.5}$$

Using (B.3), (B.4), and (B.5),

$$\begin{aligned}\mathcal{L}_{I'_k}(s) &= \exp\left(-\lambda^a_k \int_0^{2\pi}\int_r^\infty \left(1-\mathbb{E}_h\left[e^{-she^{-\alpha v^\beta}}\right]\right) v\mathrm{d}v\mathrm{d}\theta\right) \\ &= \exp\left(-2\pi\lambda^a_k \int_r^\infty \left(1-\frac{1}{1+se^{-\alpha v^\beta}}\right) v\ \mathrm{d}v\right) \end{aligned} \tag{B.6}$$

Finally, the average achievable rate per a user solely served by the k^{th} closest and given that the cell is connected is obtained as

$$\begin{aligned}\bar{R_k^M} &= \mathbb{E}\left[R_k^M\right] = \int_0^\infty \mathbb{P}\left[R_k^M > t\right] \mathrm{d}t \\ &= \int_0^\infty \mathbb{P}\left[\gamma_k > 2^{Mt} - 1\right] \mathrm{d}t. \end{aligned} \tag{B.7}$$

Using (B.2) with (B.7) yields (3.12) which completes the proof.

Proof of Corollary 3.1

Starting from (3.14) and following similar steps as in [21], we have

$$\begin{aligned}\mathcal{L}_{I_k'}(s) &= \exp\left(-2\pi\lambda_k^a \int_r^\infty \left(\frac{s}{s + e^{\alpha v^\beta}}\right) v \, \mathrm{d}v\right) \\ &\overset{(a)}{=} \exp\left(\frac{-(n+1)\pi\lambda_k^a}{\alpha^{n+1}} \int_0^{e^{-\alpha r^{\frac{2}{n+1}}}} \frac{s}{1+sx}(-\ln x)^n \mathrm{d}x\right) \\ &\overset{(b)}{=} \exp\left(\frac{-(n+1)\pi\lambda_k^a}{\alpha^{n+1}} \int_0^1 \frac{1}{w + \frac{1}{se^{-\alpha r^{\frac{2}{n+1}}}}}(-\ln w + \alpha r^{\frac{2}{n+1}})^n \mathrm{d}x\right), \end{aligned} \tag{B.8}$$

(a) is obtained by setting $x = e^{-\alpha v^\beta}$ and $\beta = \frac{2}{n+1}$ and (b) by setting $x = we^{-\alpha r^{\frac{2}{n+1}}}$. Using the binomial expansion and setting $u = -\ln w$ and $s = \zeta e^{\alpha r^{\frac{2}{n+1}}}$ with some mathematical manipulations, one can obtain (3.15) which completes the proof.

Proof of Theorem 3.2

Following the same steps in the proof of Theorem 1, the CCDF of $\gamma_h^u p$ is

$$\begin{aligned}\mathbb{P}[\gamma_h^{up} > \zeta] = & \ \mathbb{E}_{I_h^{up}, r_1}\left[\mathbb{P}\left[h > \zeta(\frac{\sigma^2}{P_0} + I_h^{up})e^{\alpha r_1^\beta} | I_h^{up}, r_1\right]\right] \\ = & \int_0^\infty e^{-\zeta\frac{\sigma^2}{P_0}e^{\alpha r^\beta}} \mathcal{L}_{I_h^{up}}(\zeta e^{\alpha r^\beta}) f_{r_1}(r)\ \mathrm{d}r, \end{aligned} \tag{B.9}$$

where the ICI experienced in the uplink HTC is $I_h^{up} = \sum_{j\in\Phi_1^a} h_j e^{-\alpha r_j^\beta}$. Note that we considered Φ_1^a not Φ_h since multiple users lying in the same Voronoi cell will share the resources orthogonally, hence, will not cause interference simultaneously. The Laplace transform of the ICI with the approximation considered in (3.20) and with replacing λ_h by λ_1^a is given by [61]

$$\begin{aligned}\mathcal{L}_{I_h^{up}}(s) &= \exp\left(-2\pi\lambda_1^a \int_{r'}^\infty \left(1 - \frac{1}{1+se^{-\alpha z^\beta}}\right) z\mathrm{d}z\right) \\ &\approx \exp\left(-2\pi\lambda_1^a \int_0^\infty (1 - e^{-\frac{12}{5}\pi\lambda_1^a z^2}) \left(\frac{se^{-\alpha z^\beta}}{1+se^{-\alpha z^\beta}}\right) z\ \mathrm{d}z\right). \end{aligned} \tag{B.10}$$

Using (B.9) and (B.10) and the fact that only a fraction ω of the bandwidth is allocated for the HTC yields (3.21) which completes the proof.

B.2 Proofs of Chapter 4

Proof of Lemma 4.2

Using (4.5) and (4.9), the CCDF on *Tier-k* is,

$$\begin{aligned}
\mathbb{P}\left(\gamma_k \geq \zeta_k\right) &= \mathbb{E}_{r_k,\Phi_{h,k}}\left[\mathbb{P}\left(\frac{h_0 r_k^{-\eta}}{I_k + \frac{\sigma^2}{P_0}} \geq \zeta_k \mid r_k, I_k\right)\right], \\
&= \mathbb{E}_{r_k,\Phi_{h,k}}\left[\mathbb{P}\left(h_0 \geq \zeta_k(I_k + \frac{\sigma^2}{P_0})r_k^{\eta} \mid r_k, I_k\right)\right], \\
&\overset{(a)}{=} \mathbb{E}_{r_k,\Phi_{h,k}}\left[\exp\left(-\zeta_k(I_k + \frac{\sigma^2}{P_0})r_k^{\eta}\right)\right], \\
&= \int_{r>0} e^{-\zeta_k \frac{\sigma^2}{P_0} r_k^{\eta}} \mathbb{E}_{I_k'}\left[\exp\Big(-\zeta_k I_k r_k^{\eta}\Big)|r_k\right] f_{r_k}(r_k)\, \mathrm{d}r_k, \\
&= \int_0^{\infty} e^{-\zeta_k \frac{\sigma^2}{P_0} r_k^{\eta}} \mathcal{L}_{I_k}(\zeta_k r_k^{\eta}) f_{r_k}(r_k)\, \mathrm{d}r_k, \qquad \text{(B.11)}
\end{aligned}$$

where $\mathbb{E}_Z[.]$ denotes expectation over random variable Z, (a) is obtained from the exponential distribution with unit mean of h_k, and $\mathcal{L}_{I_k}(s) = \mathbb{E}_{I_k}[e^{-sI_k}]$ is the Laplace transform of the normalized ICI sensed at the tagged SC on *Tier-k*. Given that $I_k = \sum_{j\in\Phi_{h,k}\setminus u_0} h_j r_j^{-\eta}$, then,

$$\begin{aligned}
\mathcal{L}_{I_k}(s) &= \mathbb{E}_{\Phi_k^h, h_j}\left[\exp\left(-s \sum_{j\in\Phi_k^h\setminus u_0} h_j r_j^{-\eta}\right)\right] \\
&\overset{(a)}{=} \mathbb{E}_{\Phi_k^h}\left[\prod_{j\in\Phi_k^h\setminus u_0} \mathbb{E}_h\Big[\exp(-s\ h\ r_j^{-\eta})\Big]\right] \\
&\overset{(b)}{=} \exp\left(-p_k\lambda_h \int_0^{2\pi}\int_{r'}^{\infty}\Big(1 - \mathbb{E}_h\left[e^{-shv^{-\eta}}\right]\Big)\, v\mathrm{d}v\mathrm{d}\theta\right) \\
&\overset{(c)}{=} \exp\left(-2\pi p_k^a \lambda_s \int_{r'}^{\infty}\left(1 - \frac{1}{1+sv^{-\eta}}\right) v\ \mathrm{d}v\right) \\
&\overset{(d)}{=} \exp\left(-\pi p_k^a \lambda_s\ s^{\frac{2}{\eta}} \int_{\frac{r'^2}{s^{2/\eta}}}^{\infty} \frac{1}{1+u^{\frac{\eta}{2}}}\mathrm{d}u\right). \qquad \text{(B.12)}
\end{aligned}$$

where (a) is obtained assuming that all small scale fading coefficients are independently and identically distributed (i.i.d.), (b) is derived using the probability generating function of an HPPP [22], (c) is obtained using (4.8) and the mean of an exponential random variable, and finally, (d) is derived by setting $u = \frac{v^2}{s^{2/\eta}}$. Substituting by (B.12) and (4.9) into (B.11) while considering an interference-limited scenario, *i.e.,* neglecting the noise power, and taking $r' \approx r_k$ for tractability which is valid in a UDN scenario ($\lambda_s > M\lambda_h$), one can obtain

$$\mathbb{P}\left(\gamma_k \geq \zeta_k\right) = \int_0^{\infty} e^{-\left(p_k^a \zeta_k^{\frac{2}{\eta}} \int\limits_{\zeta_k^{-2/\eta}}^{\infty} \frac{1}{1+u^{\frac{\eta}{2}}} \mathrm{d}u\right)\pi\lambda_s r_k^2} \times \frac{2(\pi\lambda_s)^k}{(k-1)!} r_k^{2k-1} e^{-\pi\lambda_s r_k^2}\, \mathrm{d}r_k. \tag{B.13}$$

After some mathematical manipulations and using the integration by reduction technique, (4.11) can be derived with the aid of the substitution in (4.12).

Proof of Theorem 4.1

$$\begin{aligned}
&\mathbb{P}\left(..., \gamma_i \geq \zeta_i, \gamma_j \geq \zeta_j, \gamma_k \geq \zeta_k\right) \\
&= \mathbb{E}_{...,r_i,r_j,r_k,...,\Phi_{h,i},\Phi_{h,j}\Phi_{h,k}} \Bigg[\mathbb{P}\Bigg(..., h_i \geq \zeta_i(I_i + \frac{\sigma^2}{P_0})r_i^{\eta}, h_j \geq \zeta_j(I_j + \frac{\sigma^2}{P_0})r_j^{\eta}\ , \\
&\qquad\qquad h_k \geq \zeta_k(I_k + \frac{\sigma^2}{P_0})r_k^{\eta} \mid ..., r_i, r_j, r_k, \quad ..., I_i, I_j, I_k\Bigg)\Bigg], \\
&\overset{(a)}{=} \mathbb{E}_{...,r_i,r_j,r_k,...,\Phi_{h,i},\Phi_{h,j}\Phi_{h,k}} \Bigg[... \times \exp\left(-\zeta_i(I_i + \frac{\sigma^2}{P_0})r_i^{\eta}\right) \\
&\qquad\qquad \times \exp\left(-\zeta_j(I_j + \frac{\sigma^2}{P_0})r_j^{\eta}\right) \times \exp\left(-\zeta_k(I_k + \frac{\sigma^2}{P_0})r_k^{\eta}\right)\Bigg], \\
&\overset{(b)}{=} \int_0^{\infty}\int_0^{r_k} ... \int_0^{r_3}\int_0^{r_2} ...\mathcal{L}_{I_i}(\zeta_i r_i^{\eta})\mathcal{L}_{I_j}(\zeta_j r_j^{\eta})\mathcal{L}_{I_k}(\zeta_k r_k^{\eta}) \\
&\qquad\qquad \times f_{r_1,r_2,...,r_k}(r_1, r_2, ..., r_k)\ \mathrm{d}r_1\mathrm{d}r_2...\mathrm{d}r_{k-1}\mathrm{d}r_k,
\end{aligned}$$

$$\overset{(c)}{=} \int_0^{\infty}\int_0^{r_k} ... \int_0^{r_3}\int_0^{r_2} ...e^{-Q_i(\zeta_i)\pi\lambda_s r_i^2} \times e^{-Q_j(\zeta_j)\pi\lambda_s r_j^2} \times e^{-Q_k(\zeta_k)\pi\lambda_s r_k^2} \times (2\pi\lambda_s)^k r_1 r_2 ... r_k$$

$$\times\, e^{-\pi\lambda_s r_k^2}\, \mathrm{d}r_1 \mathrm{d}r_2 ... \mathrm{d}r_{k-1} \mathrm{d}r_k. \tag{B.14}$$

(a) is obtained using the assumption that the small scale fading coefficients over the links of *Tier*-1, *Tier*-2, and up to *Tier*-k are i.i.d. and exponentially distributed with unit mean. (b) follows by assuming an interference-limited scenario where $\mathcal{L}_{I_{k'}}(s) = \mathbb{E}_{I_{k'}}[e^{-sI_{k'}}]$ is the Laplace transform of the normalized ICI sensed at the tagged *Tier*-k' SC where $k' \in \{..., i, j, k\}$. (c) is obtained by substituting the joint distance distribution of the closest k neighbours to the typical HTCU given in (4.10) and $\mathcal{L}_{I_{k'}}(\zeta_{k'} r_{k'}^{\eta}) = e^{-Q_{k'}(\zeta_{k'})\pi\lambda_s r_{k'}^2}$; $k' \in \{..., i, j, k\}$ given by (B.12) and (4.12). Finally, rearranging the different terms and using integration by parts and by reduction with some further mathematical manipulations, one can obtain (4.19).

Proof of Theorem 4.2

First, since the link association and the achievable SIR on that link are independent, then, the set of the non-associated tiers $\mathcal{M}_0$ will not affect the achievable SIRs on the remaining tiers. Accordingly, we define $\mathcal{A}_k := \gamma_k \geq 0$ to reflect a certain event, *i.e.*, $\mathbb{P}(\mathcal{A}_k) = 1,\ \forall\, k \in \mathcal{M}_0$. Using the definition of conditional probability, we follow the steps in the derivation of (B.15)

$$\mathbb{P}\left(\bigcap_{k\in\mathcal{M}_1} (\gamma_k \geq \zeta_k) \,\middle|\, \bigcap_{k\in\mathcal{M}} \mathcal{A}_k\right) = \frac{\mathbb{P}\left(\bigcap\limits_{k\in\mathcal{M}_1} (\gamma_k \geq \zeta_k),\ \bigcap\limits_{k\in\mathcal{M}} \mathcal{A}_k\right)}{\mathbb{P}\left(\bigcap\limits_{k\in\mathcal{M}} \mathcal{A}_k\right)}$$

$$= \frac{\mathbb{P}\left(\bigcap\limits_{k\in\mathcal{M}_1} (\gamma_k \geq \zeta_k),\ \bigcap\limits_{k\in\mathcal{M}_1} (\gamma_k \geq \gamma_{th}),\ \bigcap\limits_{k\in\mathcal{M}_2} (\gamma_k \leq \gamma_{th}),\ \bigcap\limits_{k\in\mathcal{M}_0} (\gamma_k \geq 0)\right)}{\mathbb{P}\left(\bigcap\limits_{k\in\mathcal{M}_1} (\gamma_k \geq \gamma_{th}),\ \bigcap\limits_{k\in\mathcal{M}_2} (\gamma_k \leq \gamma_{th}),\ \bigcap\limits_{k\in\mathcal{M}_0} (\gamma_k \geq 0)\right)},$$

$$= \frac{\mathbb{P}\left(\bigcap\limits_{k\in\mathcal{M}_1'}(\gamma_k \geq \zeta_k),\ \bigcap\limits_{k\in\mathcal{M}_1\setminus\mathcal{M}_1'}(\gamma_k \geq \gamma_{th}),\ \bigcap\limits_{k\in\mathcal{M}_2}(\gamma_k \leq \gamma_{th})\right)}{\mathbb{P}\left(\bigcap\limits_{k\in\mathcal{M}_1}(\gamma_k \geq \gamma_{th}),\ \bigcap\limits_{k\in\mathcal{M}_2}(\gamma_k \leq \gamma_{th})\right)}, \tag{B.15}$$

where the last step is obtained using the fact that $\mathbb{P}(a > b, a > c) = \mathbb{P}(a > \max(b, c))$ and that $(\gamma_k \geq 0)$ is a certain event. Combining the different terms leads directly to (4.21).

Proof of Theorem 4.3

Starting from (4.27) and using equal task division approach, *i.e.*, $\alpha_k^{\mathrm{x_i^M}} = \frac{1}{\omega^{\mathrm{x_i^M}}}\ \forall\ k \in \{0, \mathcal{M}_1^{\mathrm{x_i^M}}\}$ where $k = 0$ denotes the local processor,

$$\begin{aligned}
\mathbb{E}[t_h^{\mathrm{x_i^M}}] &= \int_0^\infty \mathbb{P}\left[t_h^{\mathrm{x_i^M}} \geq z\right] \mathrm{d}z \\
&= \int_0^\infty \left(1 - \mathbb{P}\left[t_h^{\mathrm{x_i^M}} < z\right]\right)\mathrm{d}z, \\
&= \int_0^\infty \left(1 - \mathbb{P}\left[t_0^{\mathrm{x_i^M}} < z,\ \bigcap_{k\in\mathcal{M}_1^{\mathrm{x_i^M}}}(t_k^{\mathrm{x_i^M}} < z)\right]\right)\mathrm{d}z, \\
&\overset{(a)}{=} \frac{t_{loc}}{\omega^{\mathrm{x_i^M}}} + \int_{t_{loc}/\omega^{\mathrm{x_i^M}}}^\infty \left(1 - \mathbb{P}\left[\bigcap_{k\in\mathcal{M}_1^{\mathrm{x_i^M}}}(t_k^{\mathrm{x_i^M}} < z)\right]\right)\mathrm{d}z, \\
&\overset{(b)}{=} \frac{t_{loc}}{\omega^{\mathrm{x_i^M}}} + \int_{\frac{t_{loc}}{\omega^{\mathrm{x_i^M}}}}^\infty \left(1 - \mathbb{P}\left[\bigcap_{k\in\mathcal{M}_1^{\mathrm{x_i^M}}}(\gamma_k^{\mathrm{x_i^M}} > 2^{\frac{MB_u/W}{\omega^{\mathrm{x_i^M}} z - t_{ecs}}} - 1)\right]\right)\mathrm{d}z, \\
&\overset{(c)}{=} \frac{t_{loc}}{\omega^{\mathrm{x_i^M}}} + \frac{G_M}{\omega^{\mathrm{x_i^M}}} \int_0^{\zeta_0} \left(1 - \mathbb{P}\left[\bigcap_{k\in\mathcal{M}_1^{\mathrm{x_i^M}}}(\gamma_k \geq \zeta) \,\middle|\, \bigcap_{k\in\mathcal{M}_{1\cup 2}^{\mathrm{x_i^M}}} \mathcal{A}_k^{\mathrm{x_i^M}}\right]\right) \times \frac{\mathrm{d}\zeta}{(1+\zeta)\left[\log_2(1+\zeta)\right]^2},
\end{aligned} \tag{B.16}$$

where (a) follows from the independence between the local processing time and the delay over all tiers in $\mathcal{M}_1^{\mathrm{x_i^M}}$ and the fact that $\mathbb{P}\left[t_0^{\mathrm{x_i^M}} < z\right] = 1$ whenever $z \geq t_{loc}/\omega^{\mathrm{x_i^M}}$ and zero otherwise. (b) is obtained by substituting $t_k^{\mathrm{x_i^M}}$ from (4.3) and (4.4) and setting $\alpha_k^{\mathrm{x_i^M}} = \frac{1}{\omega^{\mathrm{x_i^M}}}$. Then, considering the dependency among the achievable SINRs over all tiers and the fact

that for a given combination x_i^M, the states of all tiers $k \in \mathcal{M}$ is already known, (c) is obtained by substituting $\zeta = 2^{\frac{MB_u/W}{\omega^{\mathrm{x_i^M}} z - t_{ecs}}} - 1$, $G_M = \frac{MB_u}{W\ln(2)}$, and $\zeta_0 = 2^{\frac{G_M \ln(2)}{t_{loc} - t_{ecs}}} - 1$. Finally, using Using the results obtained in Theorem 4.2 and substituting (4.23) into (B.16) leads directly to (4.29) which completes the proof.

Proof of Lemma 4.3

The average delay experienced by the typical HTCU when the whole task (B_u, C_u) is offloaded to *Tier-k*, $k \in \mathcal{M}_1$ consists of two parts; the transmission time and the processing time. Hence,

$$\begin{aligned}
T_k^{ind} &= \mathbb{E}_{\gamma_k}\left[t_{ecs} + \frac{MB_u/W}{\log_2(1+\gamma_k|_{\gamma_k>\gamma_{th}})}\right] \\
&= t_{ecs} + \frac{MB_u}{W}\,\mathbb{E}_{\gamma_k}\left[\frac{1}{\log_2(1+\gamma_k|_{\gamma_k>\gamma_{th}})}\right] \\
&= t_{ecs} + \frac{MB_u}{W}\int_0^\infty \mathbb{P}\left[\frac{1}{\log_2(1+\gamma_k|_{\gamma_k>\gamma_{th}})} \geq z\right] \mathrm{d}z \\
&= t_{ecs} + \frac{MB_u}{W} \times \int_0^\infty \left(1 - \mathbb{P}\left[\log_2(1+\gamma_k) \geq \frac{1}{z} \;\middle|\; \gamma_k \geq \gamma_{th}\right]\right) \mathrm{d}z \\
&= t_{ecs} + \frac{MB_u}{W}\int_0^\infty \left(1 - \mathbb{P}\left[\gamma_k \geq 2^{\frac{1}{z}} - 1 \;\middle|\; \gamma_k \geq \gamma_{th}\right]\right) \mathrm{d}z. \qquad \text{(B.17)}
\end{aligned}$$

Substituting $\zeta = 2^{\frac{1}{z}} - 1$ into (B.17) and using (4.23) directly leads to (4.30).

Proof of Theorem 4.4

Following the steps in the derivation of (B.16) while using an arbitrary value of $\alpha_0^{x_i^M}$,

$$\begin{aligned}\mathbb{E}[t_{h,adp}^{x_i^M}] &= \alpha_0^{x_i^M} t_{loc} + \int_{\alpha_0^{x_i^M} t_{loc}}^{\infty} \left(1 - \mathbb{P}\left[\bigcap_{k\in\mathcal{M}_1^{x_i^M}} (t_k^{x_i^M} < z)\right]\right) \mathrm{d}z, \\ &\overset{(a)}{=} \alpha_0^{x_i^M} t_{loc} + \int_{\alpha_0^{x_i^M} t_{loc}}^{\infty} \left(1 - \mathbb{P}\left[\bigcap_{k\in\mathcal{M}_1^{x_i^M}} \left(\gamma_k^{x_i^M} > 2^{\left(\frac{MB_u/W}{z/\alpha_k^{x_i^M} - t_{ecs}}\right)} - 1\right)\right]\right) \mathrm{d}z, \\ &\overset{(b)}{=} \frac{1}{\zeta_0^{x_i^M}} + \int_0^{\zeta_0^{x_i^M}} \left(1 - \mathbb{P}\left[\bigcap_{k\in\mathcal{M}_1^{x_i^M}} (\gamma_k \geq \mathcal{Y}_k^{x_i^M}(y)) \;\middle|\; \bigcap_{k\in\mathcal{M}_{1\cup 2}^{x_i^M}} \mathcal{A}_k^{x_i^M}\right]\right) \frac{\mathrm{d}y}{y^2}. \end{aligned} \tag{B.18}$$

where (a) is obtained by substituting $t_k^{x_i^M}$ from (4.3) and (4.4) and (b) is obtained by substituting $y = 1/z$ and setting $\zeta_0^{x_i^M}$ and $\mathcal{Y}_k^{x_i^M}(y)$ as given in (4.34) and (4.35), respectively, and by considering the dependency among the achievable SINRs over all tiers and the fact that for a given combination x_i^M, the states of all tiers $k \in \mathcal{M}$ is already known.

Without loss of generality, given the task allocation fractions $\alpha_k^{x_i^M}$, $k \in \mathcal{M}_\infty^{x_i^M}$ derived from (4.31), let $\alpha_{1'}^{x_i^M} \geq \alpha_{2'}^{x_i^M} \geq ... \geq \alpha_{M_1'}^{x_i^M}$ where $M_1' = \left|\mathcal{M}_1^{x_i^M}\right|$. Then, using (4.34) and (4.35), we can deduce that $\zeta_{1'}^{x_i^M} \leq \zeta_{2'}^{x_i^M} \leq ... \leq \zeta_{M_1'}^{x_i^M}$ and $\mathcal{Y}_{1'}^{x_i^M}(y) \geq \mathcal{Y}_{2'}^{x_i^M}(y) \geq ... \geq \mathcal{Y}_{M_1'}^{x_i^M}(y)$. From Theorem 4.2, one can conclude that whenever $\mathcal{Y}_{1'}^{x_i^M}(y) < \gamma_{th}$ (*i.e.*, $y < \zeta_{1'}^{x_i^M}$), the integral in (B.18) reduces to zero. However, when $\zeta_{1'}^{x_i^M} < y < \zeta_{2'}^{x_i^M}$, only $\mathcal{Y}_{1'}^{x_i^M}(y) \geq \gamma_{th}$ will be satisfied which indicates that the set of tiers $\mathcal{M}_1'$ in (4.21) will contain only *Tier*-1′ and the integral in (B.18) will be

$$\int_{\min(\zeta_1^{x_i^M}, \zeta_0^{x_i^M})}^{\min(\zeta_{2'}^{x_i^M}, \zeta_0^{x_i^M})} \left(1 - \frac{\mathbb{P}\left[\bigcap_{k=1'}^{k=1'} \gamma_k \geq \mathcal{Y}_k^{x_i^M}(y), \bigcap_{k\in\mathcal{M}_{1\cup 2}^{x_i^M}\setminus 1'} \mathcal{A}_k^{x_i^M}\right]}{\mathbb{P}\left[\bigcap_{k\in\mathcal{M}_{1\cup 2}^{x_i^M}} \mathcal{A}_k^{x_i^M}\right]}\right) \frac{\mathrm{d}y}{y^2} \tag{B.19}$$

Following the same procedure with increasing values of y, (4.33) can be derived.

B.3 Proofs of Chapter 7

Proof of Theorem 7.1

Starting from (7.16) and following the same steps as in the proof of Lemmma 7.1., one can obtain

$$\mathbb{P}[\gamma^{AL} > \gamma_{th}] = \mathbb{E}_{\Phi_s^{LoS}}\left[\exp\left(-\frac{\gamma_{th}\ r_0^{\alpha_{LoS}^{AL}}}{P_s^0\ \beta_{LoS}^{AL}\ \mathcal{A}_{alg}}(I_0 + \sigma^2\ W_{RB}^{AL})\right)\ \ |\ \ r_0,\ I_0\right]. \tag{B.20}$$

Hence, (7.20) directly follows by integrating over the serving distance distribution given in (7.8) and given that $\mathcal{L}_{I_0}(s) = \exp(-s\ I_0)$ where $I_0 = \sum_{q\in\{alg,mis\}} I_q$ is the sensed ICI at the typical MTCD such that

$$I_q = \sum_{j\in\Phi_{s,LoS}^{a,q,alloc}\setminus SC_0} P_s^0 \beta_{LoS}^{AL} |h_j|^2 r_j^{-\alpha_{LoS}^{AL}}. \tag{B.21}$$

The laplace transform of the pdf of I_q can be obtained as follows:

$$\begin{aligned}
&\mathcal{L}_{I_q}(s) = \mathbb{E}_{I_q}\left[e^{-sI_q}\right],\\
&= \mathbb{E}_{\Phi_{s,LoS}^{a,q,alloc},|h|^2}\left[\exp\left(-s\sum_{j\in\Phi_{s,LoS}^{a,q,alloc}\setminus SC_0} P_s^0\beta_{LoS}^{AL}|h_j|^2 r_j^{-\alpha_{LoS}^{AL}}\right)\right],\\
&= \mathbb{E}_{\Phi_{s,LoS}^{a,q,alloc}}\left[\prod_{j\in\Phi_{s,LoS}^{a,q,alloc}\setminus SC_0}\mathbb{E}_{|h_j|^2}\left[\exp(-s\ P_s^0\ \beta_{LoS}^{AL}\ |h_j|^2\ r_j^{-\alpha_{LoS}^{AL}})\right]\right],\\
&\overset{(a)}{=} \mathbb{E}_{\Phi_{s,LoS}^{a,q,alloc}}\left[\prod_{j\in\Phi_{s,LoS}^{a,q,alloc}\setminus SC_0}\frac{1}{1+s\ \mathcal{A}_q\ P_s^0\ \beta_{LoS}^{AL}\ r_j^{-\alpha_{LoS}^{AL}}}\right],\\
&\overset{(b)}{=} \exp\left(-\int_0^{2\pi}\int_r^{\infty}\lambda_{s,LoS}^{a,q,alloc}(v)\ \left(1-\frac{1}{1+s\ \mathcal{A}_q\ P_s^0\ \beta_{LoS}^{AL}\ v^{-\alpha_{LoS}^{AL}}}\right) v\ \mathrm{d}v\ \mathrm{d}\theta\right)\\
&\overset{(c)}{=} \exp\left(-2\pi p_m^{alloc} p_q \lambda_s^a \int_r^{\infty}\frac{v\ e^{-\mathcal{B}_b v}}{1+\frac{1}{s\ \mathcal{A}_q\ P_s^0\ \beta_{LoS}^{AL}\ v^{-\alpha_{LoS}^{AL}}}}\ \mathrm{d}v\right)
\end{aligned}$$

$$\exp\left(-\pi\ p_m^{alloc} p_q \lambda_s^a\ (s\ \mathcal{A}_q\ P_s^0\ \beta_{LoS}^{AL})^{\frac{2}{\alpha}} \int_{\frac{r^2}{(s\ \mathcal{A}_q\ P_s^0\ \beta_{LoS}^{AL})^{2/\alpha}}}^{\infty} \frac{e^{-\mathcal{B}_b (s\ \mathcal{A}_q\ P_s^0\ \beta_{LoS}^{AL})^{1/\alpha}\sqrt{u}}}{1+u^{\frac{\alpha}{2}}} \mathrm{d}u\right), \quad \text{(B.22)}$$

where (a) is obtained from (7.1) and that for an exponentially distributed random variable $|h_q|^2$ with mean $\mathcal{A}_q$, we have $\mathbb{E}_{|h_q|^2}[e^{-\varrho|h_q|^2}] = 1/(1+\mathcal{A}_q\varrho)$, (b) is derived using the probability generating function (PGFL) of a PPP [22], where $\mathbb{E}\left[\prod_{x\in\Phi'} f(x)\right] = \exp(-\iint_{\mathbb{R}^2} \lambda'(x)(1-f(x))\mathrm{d}x)$ where λ' is the corresponding density of Φ', and (c) is obtained by substituting $\lambda_{s,LoS}^{a,q,alloc}(v) = p_m^{alloc} p_q e^{-\mathcal{B}_b v} \lambda_s^a$ where λ_s^a is the density of active SCs. Finally, setting $u = v^2 (s\ \mathcal{A}_q\ P_s^0\ \beta_{LoS}^{AL}\ v^{-\alpha_{LoS}^{AL}})^{-2/\alpha}$, we obtain (B.22). Hence, by setting $s = \frac{\gamma_{th}\ r^{\alpha_{LoS}^{AL}}}{P_s^0 \beta_{LoS}^{AL} \mathcal{A}_{alg}}$, we can obtain (7.21) which completes the proof.

Proof of Theorem 7.3

$$\begin{aligned}
\mathbb{P}[C_s^{BH} > C_{th}] &= \mathbb{E}_{\Phi_c,\Phi_s}\left[\mathbb{P}\left(\zeta_{c_s s}^{BH} > \frac{M_{c_s} W^{BH} (\sigma_w^J)^2}{P_c (N_c - M_{c_s})}\left(2^{C_{th}/W^{BH}} - 1\right) \mid M_{c_s}\right)\right], \\
&= \sum_{m_c=1,2,\ldots}^{N_c-1} \mathbb{P}\left[\zeta_{c_s s}^{BH} > \frac{m_c W^{BH} (\sigma_w^J)^2}{P_c (N_c - m_c)}\left(2^{C_{th}/W^{BH}} - 1\right)\right] \mathbb{P}\left[M_{C_s} = m_c\right], \\
&\overset{(a)}{=} \sum_{m_c=1,2,\ldots}^{N_c-1} \mathbb{P}\left[d_{c_s s}^{-\alpha^{BH}} > \frac{m_c W^{BH} (\sigma_w^J)^2}{P_c \beta^{BH} (N_c - m_c)}\left(2^{C_{th}/W^{BH}} - 1\right)\right] \mathbb{P}\left[M_{C_s} = m_c\right], \\
&= \sum_{m_c=1,2,\ldots}^{N_c-1} \mathbb{P}\left[d_{c_s s} < \left(\frac{P_c \beta^{BH} (N_c - m_c)}{m_c W^{BH} (\sigma_w^J)^2 \left(2^{C_{th}/W^{BH}} - 1\right)}\right)^{\frac{1}{\alpha^{BH}}}\right] \mathbb{P}\left[M_{C_s} = m_c\right], \qquad \text{(B.23)}
\end{aligned}$$

where (a) is obtained following that $\zeta_{c_s s}^{BH} = \beta^{BH}\ d_{c_s s}^{-\alpha^{BH}}$ is the backhaul large-scale fading channel coefficient and $d_{c_s s}$ is the distance between the CPU and the SC. Using the CDF of the distance, $F_{d_{c_s s}}(d) = 1 - e^{-\pi\lambda_c d^2}$, one can easily obtain (7.33) which completes the proof.

Bibliography

[1] S. K. Sharma and X. Wang, "Toward massive machine type communications in ultra-dense cellular iot networks: Current issues and machine learning-assisted solutions," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 426–471, 2020.

[2] "Cisco Annual Internet Report (2018–2023) White Paper," Cisco, Tech. Rep. C11-741490-01, Mar. 2020.

[3] S. Henry, A. Alsohaily, and E. S. Sousa, "5g is real: Evaluating the compliance of the 3gpp 5g new radio system with the itu imt-2020 requirements," *IEEE Access*, vol. 8, pp. 42 828–42 840, 2020.

[4] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-dense networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2522–2545, 2016.

[5] U. Siddique, H. Tabassum, E. Hossain, and D. I. Kim, "Wireless backhauling of 5g small cells: challenges and solution approaches," *IEEE Wireless Communications*, vol. 22, no. 5, pp. 22–31, 2015.

[6] G. Chopra, R. K. Jha, and S. Jain, "Novel beamforming approach for secure communication in udn to maximize secrecy rate and fairness security assessment," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 5935–5947, 2019.

[7] "White paper: Establishing the Edge: A new infrastructure model for service providers," CISCO, Tech. Rep., 2019.

[8] M. Elbayoumi, W. Hamouda, and A. Youssef, "Edge Computing and Multiple-Association in Ultra-Dense Networks: Performance Analysis," *IEEE Transactions on Communications*, vol. 70, no. 8, pp. 5098–5112, 2022.

[9] ——, "Secrecy performance in ultra-dense networks with multiple associations," in *2020 IEEE Globecom Workshops (GC Wkshps)*, 2020, pp. 1–6.

[10] ——, "A Hybrid NOMA/OMA Scheme for MTC in Ultra-Dense Networks," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–6.

[11] M. Elbayoumi, M. Kamel, W. Hamouda, and A. Youssef, "NOMA-Assisted Machine-Type Communications in UDN: State-of-the-Art and Challenges," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 1276–1304, 2020.

[12] ——, "Capacity Analysis of Downlink NOMA-Based Coexistent HTC/MTC in UDN," in *2019 International Conference on Communications, Signal Processing, and their Applications (ICCSPA)*, 2019, pp. 1–6.

[13] M. Elbayoumi, W. Hamouda, and A. Youssef, "Ergodic secrecy rate analysis of ultra-dense networks with multiple antennas," in *ICC 2021 - IEEE International Conference on Communications*, 2021, pp. 1–6.

[14] Q. Qi and X. Chen, "Wireless powered massive access for cellular internet of things with imperfect sic and nonlinear eh," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3110–3120, 2019.

[15] A. P. K. Reddy, M. S. Kumari, V. Dhanwani, A. K. Bachkaniwala, N. Kumar, K. Vasudevan, S. Selvaganapathy, S. K. Devar, P. Rathod, and V. B. James, "5g new radio key performance indicators evaluation for imt-2020 radio interface technology," *IEEE Access*, vol. 9, pp. 112 290–112 311, 2021.

[16] R. Kovalchukov, D. Moltchanov, J. Pirskanen, J. Säe, J. Numminen, Y. Koucheryavy, and M. Valkama, "Dect-2020 new radio: The next step toward 5g massive machine-type communications," *IEEE Communications Magazine*, vol. 60, no. 6, pp. 58–64, 2022.

[17] "Multi-access edge computing (MEC) 5G integration," European Telecommunications Standards Institute (ETSI), Tech. Rep. ETSI GR MEC 031 V2.1.1, Oct. 2020.

[18] M. I. Kamel, W. Hamouda, and A. M. Youssef, "Downlink coverage and average cell load of m2m and h2h in ultra-dense networks," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2017, pp. 1–5.

[19] M. Kamel, W. Hamouda, and A. Youssef, "Uplink performance of noma-based combined htc and mtc in ultradense networks," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7319–7333, 2020.

[20] J. G. Andrews, X. Zhang, G. D. Durgin, and A. K. Gupta, "Are we approaching the fundamental limits of wireless network densification?" *IEEE Communications Magazine*, vol. 54, no. 10, pp. 184–190, 2016.

[21] A. AlAmmouri, J. G. Andrews, and F. Baccelli, "Sinr and throughput of dense cellular networks with stretched exponential path loss," *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 1147–1160, 2018.

[22] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Transactions on Communications*, vol. 59, no. 11, pp. 3122–3134, 2011.

[23] M. Elbayoumi, W. Hamouda, and A. Youssef, "Multiple-association supporting htc/mtc in limited-backhaul capacity ultra-dense networks," *IEEE Transactions on Communications*, vol. 69, no. 6, pp. 4113–4127, 2021.

[24] M. Kamel, W. Hamouda, and A. Youssef, "Performance analysis of multiple association in ultra-dense networks," *IEEE Transactions on Communications*, vol. 65, no. 9, pp. 3818–3831, 2017.

[25] Y. Lin, R. Zhang, L. Yang, C. Li, and L. Hanzo, "User-centric clustering for designing ultradense networks: Architecture, objective functions, and design guidelines," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 107–114, 2019.

[26] M. I. Kamel, W. Hamouda, and A. M. Youssef, "Multiple association in ultra-dense networks," in *2016 IEEE International Conference on Communications (ICC)*, 2016, pp. 1–6.

[27] M. Agiwal, H. Kwon, S. Park, and H. Jin, "A survey on 4g-5g dual connectivity: Road to 5g implementation," *IEEE Access*, vol. 9, pp. 16 193–16 210, 2021.

[28] J. Liu, E. Ahmed, M. Shiraz, A. Gani, R. Buyya, and A. Qureshi, "Application partitioning algorithms in mobile cloud computing: Taxonomy, review and future directions," *Journal of Network and Computer Applications*, vol. 48, pp. 99–117, 2015.

[29] M. Jaber, M. A. Imran, R. Tafazolli, and A. Tukmanov, "5g backhaul challenges and emerging research directions: A survey," *IEEE Access*, vol. 4, pp. 1743–1766, 2016.

[30] G. Kwon and H. Park, "Joint user association and beamforming design for millimeter wave udn with wireless backhaul," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 12, pp. 2653–2668, 2019.

[31] R.-A. Pitaval, O. Tirkkonen, R. Wichman, K. Pajukoski, E. Lahetkangas, and E. Tirola, "Full-duplex self-backhauling for small-cell 5g networks," *IEEE Wireless Communications*, vol. 22, no. 5, pp. 83–89, 2015.

[32] A. L. Rezaabad, H. Beyranvand, J. A. Salehi, and M. Maier, "Ultra-dense 5g small cell deployment for fiber and wireless backhaul-aware infrastructures," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, pp. 12 231–12 243, 2018.

[33] Y. Teng, M. Liu, F. R. Yu, V. C. M. Leung, M. Song, and Y. Zhang, "Resource allocation for ultra-dense networks: A survey, some research issues and challenges," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2134–2168, 2019.

[34] M. Kamel, W. Hamouda, and A. Youssef, "Physical layer security in ultra-dense networks," *IEEE Wireless Communications Letters*, vol. 6, no. 5, pp. 690–693, 2017.

[35] M. Haenggi, *Stochastic geometry for wireless networks*. Cambridge University Press, 2012.

[36] H. Tabassum, E. Hossain, and J. Hossain, "Modeling and analysis of uplink non-orthogonal multiple access in large-scale cellular networks using poisson cluster processes," *IEEE Transactions on Communications*, vol. 65, no. 8, pp. 3555–3570, 2017.

[37] Y. Siriwardhana, P. Porambage, M. Liyanage, and M. Ylianttila, "A survey on mobile augmented reality with 5g mobile edge computing: Architectures, applications, and technical aspects," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 1160–1192, 2021.

[38] S. Sobhi-Givi, M. G. Shayesteh, and H. Kalbkhani, "Energy-efficient power allocation and user selection for mmwave-noma transmission in m2m communications underlaying cellular heterogeneous networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 9866–9881, 2020.

[39] N. Al-Falahy and O. Y. K. Alani, "Supporting massive m2m traffic in the internet of things using millimetre wave 5g network," in *2017 9th Computer Science and Electronic Engineering (CEEC)*, 2017, pp. 83–88.

[40] S. Han, C.-l. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5g," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 186–194, 2015.

[41] M. Polese, M. Giordani, T. Zugno, A. Roy, S. Goyal, D. Castor, and M. Zorzi, "Integrated access and backhaul in 5g mmwave networks: Potential and challenges," *IEEE Communications Magazine*, vol. 58, no. 3, pp. 62–68, 2020.

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

[43] Z. Qin, X. Yue, Y. Liu, Z. Ding, and A. Nallanathan, "User association and resource allocation in unified noma enabled heterogeneous ultra dense networks," *IEEE Communications Magazine*, vol. 56, no. 6, pp. 86–92, 2018.

[44] "Study on Non-Orthogonal Multiple Access (NOMA) for NR," 3GPP, Tech. Rep. TR 38.812, Dec. 2018.

[45] B. Makki, K. Chitti, A. Behravan, and M. Alouini, "A survey of NOMA: Current status and open research challenges," *IEEE Open J. of the Commun. Society*, vol. 1, pp. 179–189, 2020.

[46] K. S. Ali, M. Haenggi, H. ElSawy, A. Chaaban, and M. Alouini, "Downlink non-orthogonal multiple access (NOMA) in poisson networks," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1613–1628, Feb. 2019.

[47] A. Celik, M. Tsai, R. M. Radaydeh, F. S. Al-Qahtani, and M. Alouini, "Distributed user clustering and resource allocation for imperfect NOMA in heterogeneous networks," *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 7211–7227, Oct. 2019.

[48] T. Duong, X. Zhou, and H. Poor, *Trusted Communications with Physical Layer Security for 5G and Beyond*, ser. IET telecommunications series. Institution of Engineering and Technology, 2017.

[49] Y. Wu, A. Khisti, C. Xiao, G. Caire, K. Wong, and X. Gao, "A survey of physical layer security techniques for 5g wireless networks and challenges ahead," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 4, pp. 679–695, 2018.

[50] M. Elhattab and W. Hamouda, "Performance analysis for h-crans under constrained capacity fronthaul," *IEEE Networking Letters*, vol. 2, no. 2, pp. 62–66, 2020.

[51] W. Xia, J. Zhang, T. Q. S. Quek, S. Jin, and H. Zhu, "Joint optimization of fronthaul compression and bandwidth allocation in uplink h-cran with large system analysis," *IEEE Transactions on Communications*, vol. 66, no. 12, pp. 6556–6569, 2018.

[52] R. G. Stephen and R. Zhang, "Joint millimeter-wave fronthaul and ofdma resource allocation in ultra-dense cran," *IEEE Transactions on Communications*, vol. 65, no. 3, pp. 1411–1423, 2017.

[53] W. Hao, O. Muta, and H. Gacanin, "Price-based resource allocation in massive mimo h-crans with limited fronthaul capacity," *IEEE Transactions on Wireless Communications*, vol. 17, no. 11, pp. 7691–7703, 2018.

[54] J. Jang, H. J. Yang, and H. Jwa, "Resource allocation and power control in cooperative small cell networks with backhaul constraint," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 10 926–10 942, 2019.

[55] M. Polese, M. Giordani, A. Roy, S. Goyal, D. Castor, and M. Zorzi, "End-to-end simulation of integrated access and backhaul at mmwaves," in *2018 IEEE 23rd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, 2018, pp. 1–7.

[56] C. Saha, M. Afshang, and H. S. Dhillon, "Bandwidth partitioning and downlink analysis in millimeter wave integrated access and backhaul for 5g," *IEEE Transactions on Wireless Communications*, vol. 17, no. 12, pp. 8195–8210, 2018.

[57] C. Saha and H. S. Dhillon, "Millimeter wave integrated access and backhaul in 5g: Performance analysis and design insights," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 12, pp. 2669–2684, 2019.

[58] P. Parida, H. S. Dhillon, and A. F. Molisch, "Downlink performance analysis of cell-free massive mimo with finite fronthaul capacity," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, 2018, pp. 1–6.

[59] H. Masoumi and M. J. Emadi, "Performance analysis of cell-free massive mimo system with limited fronthaul capacity and hardware impairments," *IEEE Transactions on Wireless Communications*, vol. 19, no. 2, pp. 1038–1053, 2020.

[60] C. Zhu and W. Yu, "Stochastic modeling and analysis of user-centric network mimo systems," *IEEE Transactions on Communications*, vol. 66, no. 12, pp. 6176–6189, 2018.

[61] M. Haenggi, "User point processes in cellular networks," *IEEE Wireless Communications Letters*, vol. 6, no. 2, pp. 258–261, 2017.

[62] A. AlAmmouri, J. G. Andrews, and F. Baccelli, "A unified asymptotic analysis of area spectral efficiency in ultradense cellular networks," *IEEE Transactions on Information Theory*, vol. 65, no. 2, pp. 1236–1248, 2019.

[63] M. Liu, T. Song, and G. Gui, "Deep cognitive perspective: Resource allocation for noma-based heterogeneous iot with imperfect sic," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2885–2894, 2019.

[64] M. Kamel, W. Hamouda, and A. Youssef, "Uplink coverage and capacity analysis of mmtc in ultra-dense networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 746–759, 2020.

[65] S. N. Donthi and N. B. Mehta, "An accurate model for eesm and its application to analysis of cqi feedback schemes and scheduling in lte," *IEEE Transactions on Wireless Communications*, vol. 10, no. 10, pp. 3436–3448, 2011.

[66] J.-S. Ferenc and Z. Néda, "On the size distribution of poisson voronoi cells," *Physica A: Statistical Mechanics and its Applications*, vol. 385, no. 2, pp. 518 – 526, 2007.

[67] S. M. Yu and S.-L. Kim, "Downlink capacity and base station density in cellular networks," in *2013 11th International Symposium and Workshops on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, 2013, pp. 119–124.

[68] Z. Zhang, H. Sun, and R. Q. Hu, "Downlink and uplink non-orthogonal multiple access in a dense wireless network," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 12, pp. 2771–2784, 2017.

[69] Y. Liang, X. Li, and M. Haenggi, "Non-orthogonal multiple access (noma) in uplink poisson cellular networks with power control," *IEEE Transactions on Communications*, vol. 67, no. 11, pp. 8021–8036, 2019.

[70] J. Mecke, "On the relationship between the 0-cell and the typical cell of a stationary random tessellation," *Pattern Recognition*, vol. 32, no. 9, pp. 1645 – 1648, 1999.

[71] Y. Hmamouche, M. Benjillali, and S. Saoudi, "On the role of stochastic geometry in sixth generation wireless networks," in *2020 10th International Symposium on Signal, Image, Video and Communications (ISIVC)*, 2021, pp. 1–6.

[72] B. Li, M. He, W. Wu, A. K. Sangaiah, and G. Jeon, "Computation offloading algorithm for arbitrarily divisible applications in mobile edge computing environments: An OCR case," *Sustainability*, vol. 10, no. 5, 2018.

[73] M. Feng, M. Krunz, and W. Zhang, "Joint task partitioning and user association for latency minimization in mobile edge computing networks," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 8, pp. 8108–8121, 2021.

[74] K. Li, M. Tao, and Z. Chen, "Exploiting computation replication for mobile edge computing: A fundamental computation-communication tradeoff study," *IEEE Transactions on Wireless Communications*, vol. 19, no. 7, pp. 4563–4578, 2020.

[75] H. Sun, F. Zhou, and R. Q. Hu, "Joint offloading and computation energy efficiency maximization in a mobile edge computing system," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 3052–3056, 2019.

[76] F. Wang, J. Xu, X. Wang, and S. Cui, "Joint offloading and computing optimization in wireless powered mobile-edge computing systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 1784–1797, 2018.

[77] M. Mei, M. Yao, Q. Yang, M. Qin, K. S. Kwak, and R. R. Rao, "Delay analysis of mobile edge computing using poisson cluster process modeling: A stochastic network calculus perspective," *IEEE Transactions on Communications*, pp. 1–1, 2022.

[78] X. Chen, Z. Liu, Y. Chen, and Z. Li, "Mobile edge computing based task offloading and resource allocation in 5g ultra-dense networks," *IEEE Access*, vol. 7, pp. 184 172–184 182, 2019.

[79] S. Pang and S. Wang, "Joint wireless source management and task offloading in ultra-dense network," *IEEE Access*, vol. 8, pp. 52 917–52 926, 2020.

[80] H. Guo, J. Liu, and J. Zhang, "Computation offloading for multi-access mobile edge computing in ultra-dense networks," *IEEE Communications Magazine*, vol. 56, no. 8, pp. 14–19, 2018.

[81] Z. Jing, Q. Yang, M. Qin, and K. S. Kwak, "Long term max-min fairness guarantee mechanism: Adaptive task splitting and resource allocation in mec-enabled networks," in *2019 IEEE 30th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC Workshops)*, 2019, pp. 1–6.

[82] Z. Cheng, M. Min, Z. Gao, and L. Huang, "Joint task offloading and resource allocation for mobile edge computing in ultra-dense network," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–6.

[83] S.-W. Ko, K. Han, and K. Huang, "Wireless networks for mobile edge computing: Spatial modeling and latency analysis," *IEEE Transactions on Wireless Communications*, vol. 17, no. 8, pp. 5225–5240, 2018.

[84] Y. Gu, Y. Yao, C. Li, B. Xia, D. Xu, and C. Zhang, "Modeling and analysis of stochastic mobile-edge computing wireless networks," *IEEE Internet of Things Journal*, vol. 8, no. 18, pp. 14 051–14 065, 2021.

[85] H. ElSawy, A. Sultan-Salem, M.-S. Alouini, and M. Z. Win, "Modeling and analysis of cellular networks using stochastic geometry: A tutorial," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 167–203, 2017.

[86] S. Mukherjee, D. Kim, and J. Lee, "Base station coordination scheme for multi-tier ultra-dense networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 11, pp. 7317–7332, 2021.

[87] J. Yoon and G. Hwang, "Distance-based inter-cell interference coordination in small cell networks: Stochastic geometry modeling and analysis," *IEEE Transactions on Wireless Communications*, vol. 17, no. 6, pp. 4089–4103, 2018.

[88] Y. Teng, W. Sun, A. Liu, R. Yang, and V. K. N. Lau, "Mobility-aware transmit beamforming for ultra-dense networks with sparse feedback," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1968–1972, 2019.

[89] X. He, R. Jin, and H. Dai, "Physical-layer assisted secure offloading in mobile-edge computing," *IEEE Transactions on Wireless Communications*, vol. 19, no. 6, pp. 4054–4066, 2020.

[90] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017.

[91] M. Ding, D. Lopez-Pérez, G. Mao, and Z. Lin, "Performance impact of idle mode capability on dense small cell networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10 446–10 460, 2017.

[92] D. Moltchanov, "Distance distributions in random networks," *Ad Hoc Networks*, vol. 10, no. 6, pp. 1146–1166, 2012.

[93] T. D. Novlan, H. S. Dhillon, and J. G. Andrews, "Analytical modeling of uplink cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2669–2679, 2013.

[94] X. Sun, W. Yang, Y. Cai, L. Tao, Y. Liu, and Y. Huang, "Secure transmissions in wireless information and power transfer millimeter-wave ultra-dense networks," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 7, pp. 1817–1829, 2019.

[95] Y. Lin, R. Zhang, L. Yang, and L. Hanzo, "Secure user-centric clustering for energy efficient ultra-dense networks: Design and optimization," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 7, pp. 1609–1621, 2018.

[96] G. Su, M. Dai, X. Lin, B. Chen, and H. Wang, "Secure user association in ultra dense heterogeneous cellular networks with non-uniformly distributed eavesdroppers," in *2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN)*, 2019, pp. 164–168.

[97] D. Marabissi, L. Mucchi, and S. Casini, "Physical-layer security metric for user association in ultra-dense networks," in *2020 International Conference on Computing, Networking and Communications (ICNC)*, 2020, pp. 487–491.

[98] M. Forouzesh, P. Azmi, A. Kuhestani, and P. L. Yeoh, "Covert communication and secure transmission over untrusted relaying networks in the presence of multiple wardens," *IEEE Transactions on Communications*, vol. 68, no. 6, pp. 3737–3749, 2020.

[99] M. I. Kamel, W. Hamouda, and A. M. Youssef, "Coverage and capacity analysis with stretched exponential path loss in ultra-dense networks," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, 2017, pp. 1–6.

[100] M. D. Renzo, A. Guidotti, and G. E. Corazza, "Average rate of downlink heterogeneous cellular networks over generalized fading channels: A stochastic geometry approach," *IEEE Transactions on Communications*, vol. 61, no. 7, pp. 3050–3071, 2013.

[101] M. Salehi, H. Tabassum, and E. Hossain, "Accuracy of distance-based ranking of users in the analysis of NOMA systems," *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 5069–5083, Jul. 2019.

[102] T. Lv, Y. Ma, J. Zeng, and P. T. Mathiopoulos, "Millimeter-wave NOMA transmission in cellular M2M communications for internet of things," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1989–2000, Jun. 2018.

[103] Z. Zhang, Y. Hou, Q. Wang, and X. Tao, "Joint sub-carrier and transmission power allocation for mtc under power-domain noma," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2018, pp. 1–6.

[104] Z. Zhang, H. Sun, R. Q. Hu, and Y. Qian, "Stochastic geometry based performance study on 5g non-orthogonal multiple access scheme," in *2016 IEEE Global Communications Conference (GLOBECOM)*, 2016, pp. 1–6.

[105] L. Yu, J. Wu, and P. Fan, "Energy efficient designs of ultra-dense iot networks with nonideal optical front-hauls," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7934–7945, 2019.

[106] M. Rebato, J. Park, P. Popovski, E. De Carvalho, and M. Zorzi, "Stochastic geometric coverage analysis in mmwave cellular networks with realistic channel and antenna radiation models," *IEEE Transactions on Communications*, vol. 67, no. 5, pp. 3736–3752, 2019.

[107] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, 2014.

[108] M. A. Kishk and M.-S. Alouini, "Exploiting randomly located blockages for large-scale deployment of intelligent surfaces," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 4, pp. 1043–1056, 2021.

[109] M. Ibrahim, S. Elhoushy, and W. Hamouda, "Uplink Performance of MmWave-Fronthaul Cell-Free Massive MIMO Systems," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 2, pp. 1536–1548, 2022.

[110] T. Bai, R. Vaze, and R. W. Heath, "Analysis of blockage effects on urban cellular networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 9, pp. 5070–5083, 2014.

[111] S. Aditya, H. S. Dhillon, A. F. Molisch, and H. M. Behairy, "A tractable analysis of the blind spot probability in localization networks under correlated blocking," *IEEE Transactions on Wireless Communications*, vol. 17, no. 12, pp. 8150–8164, 2018.

[112] I. A. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models, design considerations, antenna constructions, and link-budget," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 2, pp. 870–913, 2018.

[113] S. Elhoushy, M. Ibrahim, and W. Hamouda, "Cell-Free Massive MIMO: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 492–523, 2022.

[114] M. E. Git Sellin *et al.*, "Enhancing 5G with microwave," *Ericsson*, Oct 2020.

[115] Albreem *et al.*, "Massive MIMO detection techniques: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3109–3132, 2019.

[116] M. Ibrahim and W. Hamouda, "Performance Analysis of Minimum Hop Count-Based Routing Techniques in Millimeter Wave Networks: A Stochastic Geometry Approach," *IEEE Transactions on Communications*, vol. 69, no. 12, pp. 8304–8318, 2021.

[117] A. M. Tulino, S. Verdú *et al.*, "Random matrix theory and wireless communications," *Foundations and Trends® in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, 2004.