Stochastic Geometry Analysis of Multiple Access, Mobility, and Learning in Cellular Networks

by

Mohammad Salehi

A thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg

Abstract

Use cases of future wireless networks (e.g. fifth-generation [5G] networks and beyond [B5G]) will have service-quality requirements including higher data rates than today's networks for enhanced mobile broadband (eMBB), minimal latency and high network availability for ultra-reliability low-latency connection (URLLC), and massive access support for machine-type communications (mMTC). Also, 5G and B5G are expected to support communications for highly mobile scenarios with applications in new vertical sectors such as unmanned aerial vehicle (UAV) and autonomous car. Therefore, 5G and B5G cellular systems require a set of new technology enablers and solutions. In this thesis, we address some of the challenges of future wireless networks. In particular, we develop novel analytical models as well as methods, which will enable us to obtain insights into the performance of large-scale cellular networks and optimize network parameters.

Non-orthogonal multiple access (NOMA) is a promising multiple access technique that enables massive connectivity and reduces the delay. We develop an analytical framework to derive the distribution of transmission success probabilities, *meta distribution*, for uplink and downlink NOMA. We also investigate the accuracy of distance-based ranking, instead of instantaneous signal power-based ranking, in the successive interference cancellation (SIC) at the NOMA receiver.

Sojourn time, the time duration that a mobile user stays within a cell, is a mobility-aware parameter that can significantly impact the performance of mobile users and it can also be exploited to improve resource allocation and mobility management methods in the network. We derive the distribution and mean of the sojourn time in multi-tier cellular networks.

Future wireless networks will exploit data-driven machine learning techniques for improving network management as well as service provisioning. Due to privacy and communication issues, learning at a centralized location (for example, at a base station) by collecting data from the mobile devices may not be always feasible. Federated learning is a machine learning setting where the centralized location trains a learning model using remote devices. Federated learning algorithms cannot be employed in real-world scenarios unless they consider unreliable and resource-constrained nature of the wireless medium. We propose a federated learning algorithm that is suitable for wireless networks.

Acknowledgments

I would first like to express my deepest gratitude to my advisor Prof. Ekram Hossain for his guidance, invaluable support, and encouragement. I am grateful for his priceless advice and great supervision. I would also like to thank him for giving me the freedom to work on the topics I liked which made my Ph.D. study a joyful journey.

I would like to thank my committee members Dr. Faouzi Bellili, Prof. Pourang Irani, and Prof Zhiguo Ding for their time and efforts. I would also like to thank them for their constructive comments and feedback that helped me improve my dissertation.

I would like to greatly thank Dr. Hina Tabassum. Hina guided me at the beginning of my Ph.D. with constructive feedback and careful reviews of my first papers. I also want to express my appreciation to all the past and present colleagues at the Wireless Communications, Networks, and Services (WiCoNS) research group for the fun times and great discussions; among them in particular Yasser, Xiao, Carlos, Arash, Vandana, Tania, Silvia, Habiba, Amal, and Angelo.

My stay at Winnipeg would not have been so memorable without the friends that I have made along the way. Thanks to Kia, Shahab, Amir, Pedram, and Mehrdad.

Lastly, but most importantly, I thank my mother, father, and brother for their continuous support and unconditional love.

To my parents

Table of Contents

List of Figures

List of Tables

Abbreviations

5G, B5G	Fifth generation, Beyond fifth generation
BS	Base station
UE	User (equipment)
PPP	Poisson point process
PCP	Poisson cluster process
PDF	Probability density function
CDF	Cumulative distribution function
CCDF	Complementary cumulative distribution function
i.i.d.	Independent and identically distributed
PGFL	Probability generating functional
CSP	Conditional success probability
SINR	Signal-to-interference-plus-noise ratio
SIR	Signal-to-interference ratio
OMA	Orthogonal multiple access
NOMA	Non-orthogonal multiple access
SIC	Successive interference cancellation
ISP	Instantaneous signal power
MSP	Mean signal power
CSI	Channel state information
FL	Federated learning
SGD	Stochastic gradient descent

Chapter 1

Introduction

1.1 Introduction

Performance targets and standard services of each generation of cellular systems are defined based on new applications and technologies [1]. To understand the motivation behind deployment of fifth generation (5G) and beyond (B5G) cellular systems, consider the following example from [2]:

Imagine "a typical morning office commute in a 5G-connected car just a few years down the road. The vehicle is constantly exchanging position, behavior, and system status information with nearby vehicles, the surrounding highway infrastructure, and traffic control centers. Doing so in a fast and reliable manner augments the car's awareness of its surroundings and allows the driver to turn the steering, accelerating, and braking functions over to the car's semi-autonomous driving system. He can now focus on the morning's first conference call.

The driver's team is trying to find the root cause of a turbine malfunction. He puts on his augmented reality (AR) set, and a wireless 4K video feed of an airplane turbine overlaid with sensor data and gauge readings fills his screen. Collaborating in real time with a group of engineers in three different countries, the team guides a technician to isolate one of the components and recommends a troubleshooting procedure.

A few minutes later, when his intelligent-highway exit comes up, the driver takes back control of the car, switches over to a low-bandwidth voice-only connection, and drives into work. The car guides him to the closest available parking spot with an electric charging station. The parking sensor at that spot detects his car and updates the parking availability information on the network. When he plugs in the car to charge, the charging terminal establishes a low-data-rate connection to verify his account and process payment."

Future use cases of 5G request diverse service-quality requirements including higher data rates than to-
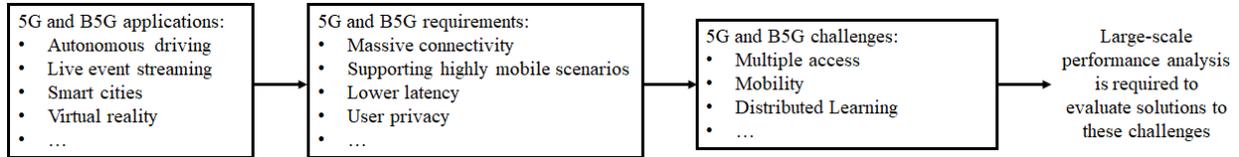
Figure 1.1: Overview

day's networks for enhanced mobile broadband (eMBB), minimal latency and high network availability for ultra-reliability low-latency connection (URLLC), and massive access support for machine-type communications (mMTC) [3]. These service classes are expected to morph and merge in B5G wireless networks [1]. As a result, 5G and B5G cellular systems require a set of new technology enablers and solutions to accomplish their goals. Performance targets in 5G and B5G can be achieved through 1) redesigning the multiple access format, 2) integrating aerial and terrestrial networks, 3) distributing machine learning at network edge, 4) communicating with large intelligent surfaces, 5) moving toward higher frequency bands, etc [1, 4, 5].

Before deploying these solutions in large-scale systems, we must identify corresponding challenges and performance gains/losses. In this regard, in this dissertation, we study some of the existing challenges in large-scale deployment of 5G and beyond cellular wireless networks.

1.2 Road Map

5G and B5G wireless networks must have sufficient flexibility and capability to support existing and emerging applications including autonomous driving, smart city, live event streaming, and virtual reality. It is not possible to achieve heterogeneous sets of requirements of this broad range of applications by a one-size-fits-all solution, i.e., there is no single solution that supports the super-set of all sets of requirements. In this regard, for 5G and B5G networks we need

- a new **multiple access** format to address the increasing demand for wireless communications,

- better **mobility** management techniques to enable applications in highly mobile scenarios such as autonomous vehicles and drones,

- efficient use of the available data, through **distributed learning**, for improving network performance and service provisioning.

Thus, multiple access, mobility, and distributed learning are challenging for future networks in the sense that current solutions to these challenges are not sufficient to support future requirements. To ensure that proposed solutions for multiple access, mobility, and distributed learning are capable of accomplishing

2

their performance targets in 5G and B5G, we must study the large-scale performance of each (Fig. 1.1). Accordingly, in this dissertation, we

- analyze performance of NOMA, a promising multiple access technique for future wireless networks, in large-scale cellular networks to enable massive connectivity,

- derive sojourn time distribution, a mobility-related performance metric, for large-scale networks to understand effect of mobility and enable better mobility management for highly mobile scenarios,

- design a distributed learning algorithm for large-scale cellular networks to efficiently use the available data at user devices.

To this end, in this chapter, we briefly explain large-scale challenges and our contributions related to multiple access, mobility, and distributed learning. In the next chapter, we introduce large-scale cellular networks and provide mathematical preliminaries, and, in the subsequent chapters, we use tools from stochastic geometry to tackle the challenges.

1.3   Multiple Access in Future Wireless Networks

As mentioned earlier, next generations of wireless networks must provide service to a massive number of users, support rate-hungry applications, guarantee low-latency communication, and provide reliable connectivity. Effective multiple access technique facilitates meeting these requirements. Adopted multiple access techniques in the previous generations of wireless networks, from the first generation (1G) to the fourth generation (4G), have one common theme—allocate orthogonal resource blocks to different users [6]. For example, time division multiple access (TDMA) allocates different time slots to different users, or frequency division multiple access (FDMA) allocates different frequency channels. However, supporting billions of devices and acquiring low-latency communication for emerging massive wireless networks necessitate efficient user access mechanisms that can potentially serve multiple devices in the same resource block [7].

Non-orthogonal multiple access (NOMA) has been recognized as a promising multi-user channel access technique that enables massive connectivity while reducing the transmission delay of the devices [7]. Contrary to traditional orthogonal multiple access (OMA), such as time division multiple access (TDMA), frequency division multiple access (FDMA), and code division multiple access (CDMA), the key idea of NOMA is to serve multiple users, termed NOMA cluster users, in the same channel simultaneously. The concurrent transmissions in NOMA reduce the waiting time of the devices while saving network resources. Of course, this can be achieved at the expense of additional interference and decoding complexity at the receivers.

3

(a) Downlink.

(b) Uplink.
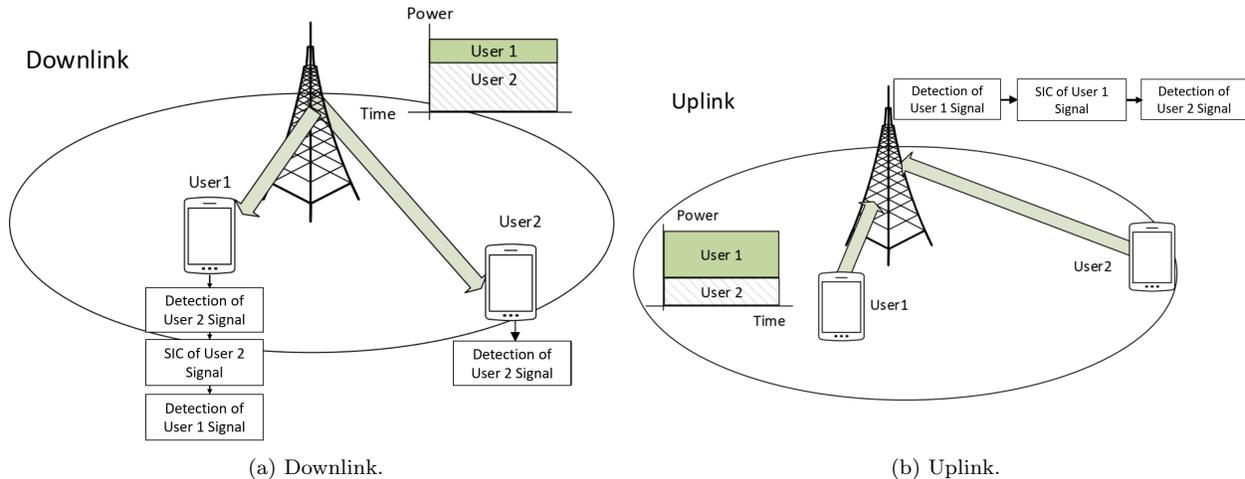
Figure 1.2: Two-user NOMA.

NOMA uses superposition coding at the transmitters, and exploits successive interference cancellation (SIC) at the receiver(s) to mitigate interference [8, 9] as shown in Fig. 1.2.

Recently, performance analysis of NOMA-based wireless networks has attracted significant research interest. The existing studies contribute mainly toward understanding the average performance of users considering a single NOMA cell/cluster [8, 10, 11, 12, 13, 14]. For instance, the performance of a single-cell downlink NOMA system with randomly located users was first studied in [10]. In particular, the signal-to-interference-plus-noise ratio (SINR) outage probability and the ergodic capacity were derived for a user at rank m in terms of distance.

The aforementioned research studies ignore the impact of inter-cell interference (interference from transmissions in other cells) which can significantly limit the performance of NOMA in massive wireless networks. Very recently, some of the research works have considered performance characterization of large-scale NOMA systems using stochastic geometry tools [15, 16, 17, 18, 19].

Problem 1: To derive the analytical results for uplink of cellular networks with PPP (Poisson point process) distributed base stations (BSs)[1], it was assumed that uplink inter-cell interferers form a homogeneous PPP which is inaccurate since there is no inter-cell interferer in the desired cell, where the desired user lies in [20]. Thus, the effective inter-cell interference field as observed from the desired BS can be modeled more accurately as an inhomogeneous PPP as demonstrated in [21, 22].

Contribution 1: Since the actual point process for the spatial locations of the inter-cell interferers is unknown, we propose two models for this point process based on the pair correlation between interferers and a typical BS. We demonstrate the accuracy of the proposed point processes by comparing the second

---

[1]This model is explained in the next chapter.

moment measure of each process with that of the original process obtained via simulations.

As explained earlier, performance of NOMA in both uplink and downlink depends on the SIC which relies on the ranking of the users in each NOMA cluster [15]. In particular, downlink intra-cell interference (interference from transmissions within the desired cell) received at a given user in NOMA depends on the power allocation factors of users in the cluster. These power allocation factors are designed according to the ranking of users' transmission links quality. For example, users with stronger links have smaller power allocations and vice versa. On the other hand, in uplink NOMA to apply SIC, BS successively decodes and cancels the messages of strong channel users prior to decoding the signals of weak channel users [15]. Therefore, the intra-cell interference encountered by any user depends on the instantaneous received signal powers (which includes short-term fading) of users in the NOMA cluster.

Problem 2: To avoid analytical complexity (in theory) and overcome implementation complexity, mean signal power- (distance-) based ranking is typically considered to be appropriate for ordering users in a NOMA cluster. Although this method simplifies the analysis and provides tractable results, its validation (i.e., *accuracy*) has not been studied yet. The distance-based ordering may not always be accurate, especially in a dynamic multi-path fading environment, where a near user can experience severe fading and a far user can observe weak fading. Since the ranking of users in a NOMA system has a direct impact on the system performance (e.g., coverage probability) analysis, it is crucial to quantify the impact of distance-based ranking in various environments and to identify the scenarios where this ranking is accurate (i.e., provides system performance close to that achievable with full CSI-based user ranking).

Contribution 2: We characterize the accuracy of analyzing the performance of a NOMA system where users are ranked according to their distances (or, equivalently, mean signal powers) instead of instantaneous signal powers, i.e., product of their distance-based path-loss and fading channel. In particular, we derive the probability of the event when distance-based ranking yields same results as instantaneous signal power-based ranking, which is referred to as the *accuracy probability*.

1.4   Mobility in Future Wireless Networks

The next generations of cellular wireless networks are expected to support communications for highly mobile users and devices [23] with applications in new vertical sectors. For instance, in 5G and B5G wireless communication systems, drones can be exploited as aerial base stations to assist terrestrial communications, or drones with their own missions could be connected to terrestrial networks as new aerial users [24]. The mobility of drones allows flexible network reconfiguration to adapt to dynamic traffic and channel conditions. It also enables a wide range of key applications that include military, surveillance and monitoring,

5

telecommunications, delivery of medical supplies, and rescue operations [24].

Therefore, addressing the mobility related challenges is necessary for the development of the next generation cellular networks. Impact of user/device mobility on its performance in cellular networks can be measured through mobility-aware performance metrics such as handoff rate, handoff probability, and sojourn time [23]. Sojourn time (or dwell time), time duration that a mobile user stays within a cell, is a key network parameter which allows studying other important network parameters such as channel occupancy time, new call and handoff call dropping probabilities [25]. Therefore, it is imperative to incorporate the sojourn time distribution in resource allocation and mobility management for improving the network performance.

Problem 3: To analyze the sojourn time distribution in multi-tier scenarios with PPP distributed BSs, the existing works either assume that the mobile user is always associated to only one of the tiers [26], or only focus on the small tier (in two-tier scenarios) [27, 28]. For both the cases, the results are no different from the single-tier scenarios [29]. In single-tier networks with maximum averaged received power association (nearest BS association), (Voronoi) cells are convex; however, in multi-tier networks with maximum biased averaged received power association, cells may not be convex depending on the transmission power and bias factor of each tier. Therefore, analysis of sojourn time of multi-tier cellular networks is more complicated compared to the single-tier networks.

Contribution 3: We derive the distribution and mean of the sojourn time for multi-tier cellular networks, and study its relation with other mobility-related performance metrics.

1.5   Distributed Learning in Wireless Networks

With growth in the number of smart devices and advancements in their hardware, in recent years, data-driven machine learning techniques have drawn significant attention. In fact, machine learning techniques analyze the data and obtain useful information for the detection, classification, and prediction of future events [30]. For example, app Waze suggests the optimal route to drivers by collecting a large number of data from users including every road driven [31], or in edge computing users data can be used to determine popular contents for caching. However, due to privacy and communication issues, it is not always possible to collect the data at a centralized location. Federated learning (FL) is a machine learning setting where the centralized location trains a learning model over remote devices.

In FL, a learning model is trained over remote devices under the control of a centralized location, called server [32, 33, 34]. Specifically, in federated learning, the server broadcasts the global model parameters to the remote devices. Each remote device uses its local dataset to update the global model, and then transmits

the updated local model to the server. After aggregating the local models, the server updates the global model and repeats the whole procedure.

Problem 4: Training federated learning models in a wireless networking environment requires devices and the server exchange information via wireless transmissions. Due to the limited number of available resource blocks (such as bandwidth) for communication, full device participation at each round of communication is not possible. Moreover, in reality, transmission success probability is different for different devices, even when they all have the same hardware. Specifically, transmission success probability for devices that are located closer to the server is higher than far devices. Thus, without considering this issue at the time of updating the global model, the updated global model will be biased towards local models of devices that are closer to the server. Existing works [35, 36, 37, 38, 39, 40] assume that transmissions are always successful, which is not correct in reality.

Contribution 4: We propose a federated learning algorithm that is tailored for unreliable and resource-constrained wireless cellular networks, where success probability varies across different devices.

1.6 Organization of the Thesis

Chapter 2 provides mathematical preliminaries and introduces Poisson cellular network. As discussed earlier, in this thesis, we study multiple access, mobility, and learning in Poisson cellular networks.

In this regard, in Chapters 3 and 4, we study the large-scale performance of NOMA technique. Specifically, we focus on **Problem 1** in Chapter 3 and **Problem 2** in Chapter 4. These chapters are based on the following published papers:

[1] **M. Salehi**, H. Tabassum, and E. Hossain, "Meta distribution of the SIR in large-scale uplink and downlink NOMA networks," *IEEE Trans. on Commun.*, vol. 67, no. 4, pp. 3009-3025, Apr. 2019.

[2] **M. Salehi**, H. Tabassum, and E. Hossain, "Accuracy of distance-based ranking of users in the analysis of NOMA systems," *IEEE Trans. on Commun.*, vol. 67, no. 7, pp. 5069-5083, July 2019.

[3] **M. Salehi** and E. Hossain, "On coverage probability in uplink NOMA with instantaneous signal power-based user ranking," *IEEE Wireless Commun. Letters*, vol. 8, no. 6, pp. 1683-1687, Dec. 2019.

In Chapter 5, we derive the mean and distribution of the sojourn time in multi-tier cellular networks. This chapter is based on the following published papers:

[4] H. Tabassum, **M. Salehi**, and E. Hossain, "Fundamentals of mobility-aware performance characterization of cellular networks: A tutorial," *IEEE Commun. Surveys and Tutorials*, vol. 21, no. 3, pp. 2288-2308, Third Quarter, 2019.

[5] **M. Salehi** and E. Hossain, "Stochastic geometry analysis of sojourn time in multi-tier cellular networks," *IEEE Trans. on Wireless Commun.*, DOI: 10.1109/TWC.2020.3036894.

[6] **M. Salehi** and E. Hossain, "Handover rate and sojourn time analysis in mobile drone-assisted cellular networks," *IEEE Wireless Commun. Letters*, DOI: 10.1109/LWC.2020.3032596.

In Chapter 6, we work on **Problem 4**, where we propose an FL algorithm and study its convergence in unreliable and resource-constrained cellular wireless networks. This chapter is based on the following submitted manuscript:

[7] **M. Salehi** and E. Hossain, "Federated learning in unreliable and resource-constrained cellular wireless networks," *IEEE Trans. on Commun.*, to appear.

Finally, the dissertation concludes with Chapter 7, which summarizes the key contributions and discusses promising future directions of research.

Chapter 2

System Model and Mathematical Preliminaries

As discussed in the previous chapter, in this dissertation, we study multiple access, mobility, and learning in Poisson cellular networks. In this regard, in this chapter, we first provide mathematical preliminaries, then introduce the network model, and finally present performance metrics that we use to study multiple access, mobility, and learning.

2.1 Mathematical Preliminaries

Stochastic geometry studies random spatial patterns [41]. The basic ingredients of geometry are points. Point process theory is the main sub-field of stochastic geometry and studies random point patterns [42, 43].

Following imaginary scenario from [42] helps to better understand point processes: A simple example of a random variable describes the outcome of a roll of a die. Here the random variable takes on the values 1 to 6, and when the die is fair the probability for each of the six values is $\frac{1}{6}$. Now assume that it is possible to construct a die with infinitely many sides, with a (two-dimensional) point pattern on each of the sides. Every time the die is thrown, a point pattern is generated. This die represents the point process Φ, and a point pattern $\varphi = \Phi(w)$ is assigned to each sample point w represented by a roll of die. Now assume somebody is observing the independent repetitive rolling of the die. This person sees different point patterns and, in particular, fluctuating values of the number of points in some fixed, deterministic subset B of the plane. In other words, every throw of the 'point pattern die' produces a different realisation of the random variable 'number of points in B'.

Therefore, there are two ways to interpret a point process Φ [41, 21]. In Random set formalism, $\Phi = \{x_1, x_2, ...\}$ is regarded as a random set of points in space (usually the Euclidean space $\mathbb{R}^d$.), where each element x_i is a random variable. In Random measure formalism, $\Phi(B)$ is a random variable that assumes values from the non-negative integers and denotes the number of points falling in set $B \subset \mathbb{R}^d$.

Definition 2.1 (Intensity Measure [41]). *For any point process Φ, the intensity measure $\Lambda(B)$ is the mean number of points in B, i.e.,*

$$\Lambda(B) = \mathbb{E}[\Phi(B)],$$

where the expectation is with respect to the point process Φ.

Definition 2.2 (Stationarity [41, 42]). *For a point process $\Phi = \{x_i\}$ to be stationary, $\Phi = \{x_i\}$ and $\Phi = \{x_i + x\}$ must have the same distribution for all $x \in \mathbb{R}^d$, i.e., Φ is stationary if its distribution is translation-invariant.*

Definition 2.3 (Isotropy [41, 42]). *Φ is isotropic if its distribution is invariant with respect to rotations about the origin.*

For stationary point processes, the intensity measure is just a multiple of the Lebesgue measure, i.e.,

$$\Lambda(B) = \lambda|B|,$$

where λ is called the intensity of Φ, and $|B|$ is the Lebesgue measure of B. Thus, $|B|$ denotes area of the set B in two-dimensional space and volume of the set B in three-dimensional space.

Definition 2.4 (Proability generating functional (PGFL) [21]). *PGFL of point process Φ with respect to a function $f : \mathbb{R}^d \to \mathbb{R}^+$ is*

$$G(f) = \mathbb{E}\left[\prod_{x \in \Phi} f(x)\right].$$

Poisson Point Process (PPP)

Poisson point process (PPP) Φ with intensity measure Λ is characterised by two properties [21, 42, 43]:

- *Poisson distribution of point counts*: $\Phi(B)$ is a Poisson random variable with mean $\Lambda(B)$ for every set B. Therefore,
$$\mathbb{P}(\Phi(B) = k) = \frac{(\Lambda(B))^k}{k!} \exp(-\Lambda(B)).$$

- *Independent scattering*: For all disjoint sets $B_1, B_2, ..., B_m$, random variables $\Phi(B_1), \Phi(B_2), ..., \Phi(B_m)$ are independent.

Homogeneous PPP is a special case of PPP where $\Lambda(B) = \lambda|B|$, for any set B [41]. Homogeneous PPP is stationary and isotropic.

Definition 2.5 (PGFL of PPP). *For PPP Φ with intensity measure Λ, PGFL with respect to f is given as*

$$G(f) = \mathbb{E}\left[\prod_{x \in \Phi} f(x)\right] = \exp\left\{-\int_{\mathbb{R}^d} (1 - f(x))\Lambda(\mathrm{d}x)\right\}.$$

When Φ is a homogeneous PPP with intensity λ, we can further simplify the above equation as

$$G(f) = \mathbb{E}\left[\prod_{x \in \Phi} f(x)\right] = \exp\left\{-\lambda \int_{\mathbb{R}^d} (1 - f(x))\mathrm{d}x\right\}.$$

2.2 Network Model

Prior to developing stochastic geometry for modeling and analysis of cellular wireless networks, system engineers and researchers used regular hexagonal or square lattice for modeling. However, for such models tractable expressions for average performance are unavailable and simulations are also time-consuming. On the other hand, modeling networks using tools from stochastic geometry yields closed-form expressions which not only circumvents the need for simulation but also provides useful insights for system design [44].

To model the cellular networks, we assume that base stations (BSs) are spatially distributed following a homogeneous PPP Φ_{B} with intensity λ_{b}[1]. Users are located according to a homogeneous PPP Φ_{U} with intensity λ_{u}. This model is in direct agreement with the 3rd generation partnership project (3GPP) models [45], and has been widely used in the literature ([46, 47] and references therein).

Real world deployment of BSs are neither perfectly regular (as in hexagonal grid) nor completely random (as in PPP). Indeed, the actual deployment of BSs lies between perfectly regular and completely random. Thus, perfectly regular (such as hexagonal grid) and completely random (such as PPP) spatial models provide bounds for the actual network performance [21]. Due to the analytical tractability of PPP models, existing literature mostly resort to PPP to model the cellular networks. Recently, it has been observed that performance of a variety of cellular network models and transmission techniques look very similar in shape. In fact, they can be well approximated by a simple horizontal shift of the most tractable model, the Poisson point process (PPP) [48]. Moreover, in many circumstances, cellular networks' users take independent locations which motivates using PPP to model the spatial distribution of users.

A realization of a cellular network over a $100\,\mathrm{km} \times 100\,\mathrm{km}$ area with $\lambda_{\mathrm{b}} = 0.004\,\mathrm{points/km}^2$ and $\lambda_{\mathrm{u}} = 0.02\,\mathrm{points/km}^2$ is shown in Fig. 2.1. Each user in the network is connected to its nearest BS. Therefore,

---

[1]In this chapter, we only focus on single-tier networks.
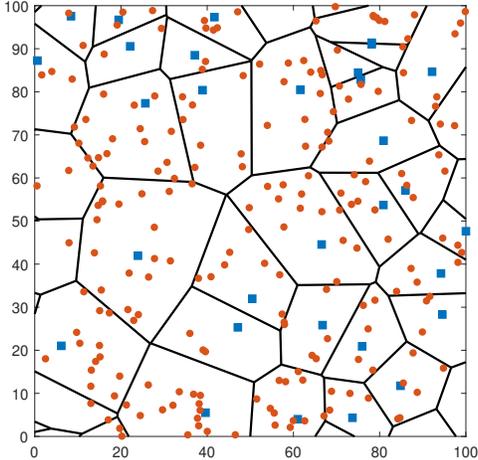
Figure 2.1: A realization of a cellular network over a $100\,\mathrm{km} \times 100\,\mathrm{km}$ area with $\lambda_\mathrm{b} = 0.004\,\mathrm{points/km}^2$ and $\lambda_\mathrm{u} = 0.02\,\mathrm{points/km}^2$. BSs are represented by blue squares, users are shown by brown circles, and cells' boundaries are illustrated by black lines.

each BS serves the users that are located in its Voronoi cell. For a BS at x, Voronoi cell $V(x)$ is given as [41]

$$V(x) = \left\{ y \in \mathbb{R}^2 : \|y - x\| \leq \|y - z\|, \quad \forall z \in \Phi_\mathrm{b} \setminus \{x\} \right\}{}^2.$$

2.3 SINR Distribution

Because of the stationarity of the homogeneous PPP, we can condition on having a receiver at the origin which is called the typical receiver. In downlink communications, the typical receiver belongs to Φ_U, while it belongs to Φ_B in uplink. For such a set-up, the concepts of success/coverage probaility, CSP, and the meta distribution along with their evaluation methods are defined in the following to provide a preliminary mathematical background to readers. We denote the transmitter point process by Φ. $\Phi \equiv \Phi_\mathrm{B}$ in downlink, and $\Phi \equiv \Phi_\mathrm{U}$ in uplink.

Definition 2.6 (Success Probability). *Success probability is the complementary cumulative distribution function (CCDF) of the signal-to-interference-plus-noise ratio (SINR), i.e., $\mathbb{P}(\mathrm{SINR} > \theta)$.*

Definition 2.7 (Conditional Success Probability (CSP) [49]). *Given the location of the transmitters Φ and conditioned on the desired transmitter to be active (denoted by tx), CSP is defined as:*

$$P_\mathrm{s}(\theta) \triangleq \mathbb{P}(\mathrm{SINR} > \theta \mid \Phi, \mathrm{tx}), \tag{2.1}$$

---

[2]In this dissertation, we are interested in 2-dimensional Euclidean space.

where θ is the desired SINR and the b-th moment of $P_\mathrm{s}(\theta)$ is given by $M_b(\theta) = \mathbb{E}_\Phi \left[P_\mathrm{s}(\theta)^b \right]$. Therefore, success probability is the first moment of the CSP.

Using the indicator function $\mathbf{1}(.)$, success probability can be written as

$$\mathbb{P}(\mathrm{SINR} > \theta) = \mathbb{E}\left[\mathbf{1}\left(\mathrm{SINR} > \theta \right) \right],$$

where the expectation is over time and space. However, conditional success probability is as

$$\mathbb{P}(\mathrm{SINR} > \theta \mid \Phi, \mathrm{tx}) = \mathbb{E}\left[\mathbf{1}\left(\mathrm{SINR} > \theta \right) \mid \Phi, \mathrm{tx} \right],$$

i.e., conditional success probability is obtained by only averaging with respect to time (while keeping spatial random variable Φ fixed).

Definition 2.8 (Meta Distribution of CSP)**.** *Meta distribution is the CCDF of $P_\mathrm{s}(\theta)$, i.e.,*

$$\bar{F}_{P_\mathrm{s}(\theta)}(x) \triangleq \mathbb{P}(P_\mathrm{s}(\theta) > x), \qquad x \in [0, 1]. \tag{2.2}$$

Since Φ_B and Φ_U are ergodic, the meta distribution can be interpreted as the fraction of active users whose success probabilities are more than x in each realization. In [49], an exact expression along with an approximation and simple bounds for the meta distribution are provided. A summary of these results is given below.

- *Exact meta distribution of CSP:* To derive the exact meta distribution, we first need to derive imaginary moments $M_{jt}(\theta) = \mathbb{E}_\Phi \left[P_\mathrm{s}(\theta)^{jt} \right]$, where $j = \sqrt{-1}$ and $t \in \mathbb{R}^+$. Then using the Gil-Pelaez theorem [50], the exact meta distribution is given as follows:

$$\bar{F}_{P_\mathrm{s}(\theta)}(x) = \frac{1}{2} + \frac{1}{\pi} \int\limits_0^\infty \frac{\Im\left(e^{-jt \log x} M_{jt}(\theta) \right)}{t} \mathrm{d}t, \qquad x \in [0, 1], \tag{2.3}$$

  where $\Im(s)$ gives the imaginary part of s.

- *Approximate meta distribution of CSP:* A simple approximation of the meta distribution is provided by using the beta distribution. In this approach, we need to derive the first moment $M_1(\theta)$ and the second moment $M_2(\theta)$ of $P_\mathrm{s}(\theta)$ and match them with the first and second moments of the beta distribution, i.e.,

$$\bar{F}_{P_\mathrm{s}(\theta)}(x) \approx 1 - I_x \left(\frac{M_1(\theta)\beta}{1 - M_1(\theta)}, \beta \right), \qquad x \in [0, 1], \tag{2.4}$$

13

where $\beta = \frac{(M_1(\theta) - M_2(\theta))(1 - M_1(\theta))}{(M_2(\theta) - M_1(\theta)^2)}$, $I_x(a, b)$ is the regularized incomplete Beta function, and $B(a, b)$ is the Beta function. The beta distribution [49, 20, 51, 52] and the generalized beta distribution [53] have been shown to match the exact meta distribution.

- *Bounds on the meta distribution* are also presented in [49, Corollary 4]. For the Markov's bound, we can use any moment of $(1 - P_s(\theta))$ and $P_s(\theta)$. For the Chebyshev's bound, we need the mean $(M_1(\theta))$ and variance $(M_2(\theta) - M_1(\theta)^2)$ of CSP. For the Paley-Zygmund (or Cauchy-Schwarz) bound, we simply need the first moment $M_1(\theta)$.

For a given realization of transmitters Φ, the transmission success events are obtained by averaging over the fading channels and are thus i.i.d. over time for i.i.d. fading. The *local delay* (defined as the number of transmission attempts until a packet is successfully received [54, 55]), is thus geometrically distributed with parameter $P_s(\theta)$. For the rest of this chapter and in the following chapters, to simplify notation, we omit θ from $P_s(\theta)$ and $M_b(\theta)$. Thus, $P_s \triangleq P_s(\theta)$ and $M_b \triangleq M_b(\theta)$.

Definition 2.9 (Distribution of the Local Delay). *For a given realization, local delay, L, follows a geometric distribution with parameter P_s given in* **Definition 2.7**, *i.e.,*

$$\mathbb{P}(L = k \mid \Phi) = (1 - P_s)^{k-1} P_s, \qquad k \in \mathbb{N}. \tag{2.5}$$

Therefore, the mean local delay is given by $\mathbb{E}[L] = \mathbb{E}_\Phi[\mathbb{E}[L \mid \Phi]] = \mathbb{E}_\Phi\left[\frac{1}{P_s}\right] = M_{-1}$ *and the variance of the local delay is* $\mathbb{E}[L^2] - \mathbb{E}[L]^2 = \mathbb{E}_\Phi[\mathbb{E}[L^2 \mid \Phi]] - M_{-1}^2 = 2M_{-2} - M_{-1} - M_{-1}^2$,

For ergodic point processes, now we are able to answer the question "what fraction of users successfully receive their desired signals in at most k time slots with probabilities larger than x?". We can answer this question by deriving the following:

$$\mathbb{P}(\mathbb{P}(L \leq k \mid \Phi) > x) = \mathbb{P}\left(1 - (1 - P_s)^k > x\right) = \bar{F}_{P_s}(1 - (1 - x)^{1/k}), \tag{2.6}$$

where $\bar{F}_{P_s}(.)$ is the meta distribution defined in (2.2). Based on (2.6), the meta distribution also reveals the distribution of the CSP for any number of retransmissions.

Example: To better understand conditional success probability (CSP), success probability, mean local delay, and the meta distribution consider the example shown in Fig. 2.2. For each user in Fig. 2.2, CSP is written above its link and is obtained by averaging over fading (time averaging). The success probability in this example is $(0.1 + 0.4 + 0.4 + 0.9)/4 = 0.45$, i.e., the success probability is the average (first moment) of the conditional success probabilities over different users (space). For device with conditional success probability
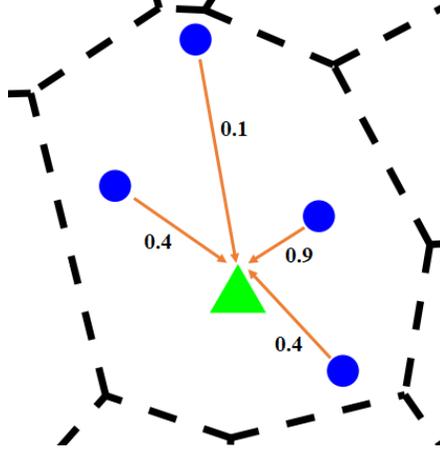
Figure 2.2: A typical cell with four users. Conditional success probability is written above each link.

0.1, local delay is $\frac{1}{0.1} = 10$ since on average it transmits successfully once in every ten attempts. Similarly, local delay for users with conditional success probability 0.4 is $\frac{1}{0.4}$ and it is $\frac{1}{0.9}$ for the user with conditional success probability 0.9. Therefore, mean local delay is $\left(\frac{1}{0.1} + \frac{1}{0.4} + \frac{1}{0.4} + \frac{1}{0.9}\right)/4 \approx 4$, i.e., mean local delay is the average of $\frac{1}{\text{CSP}}$ (first negative moment of CSP) over different users (space). Finally, meta distribution gives information about the distribution of CSPs. From the meta distribution, we can understand that for $\frac{1}{4}$ of users CSP is 0.1, for half of the users CSP is 0.4, and for the rest of the users it is 0.9.

Next, to understand the importance of meta distribution, consider the following example.

Example: Consider two networks. Assume that, in network A, the CSP of half of the users is 0 while that of the other half is 1. In network B, the CSP is 0.5 for all the users. For these two networks, the standard (average) success probability is the same and is equal to 0.5, while they behave completely different in terms of delay. In network A, half of the users have a successful transmission after their first attempt while the transmission delay is infinity for the other half of the users (i.e., the mean local delay is infinity when CSP of users is 0 or close to 0). In network B, the number of transmission attempts until the first successful reception is almost surely finite for all the users, and for this network, the average number of transmission attempts for successful transmission is 2. Therefore, comparing two networks simply in terms of their average success probabilities is not accurate since average success probability is not enough to capture a network performance. On the other hand, meta distribution is a fine-grained performance metric that provides detailed information about distribution of the conditional success probability. In fact, from the meta distribution we can calculate average success probability and mean local delay, this is similar to calculating moments of a random variable using its distribution function.

Example: With $x = 0.95$, $\bar{F}_{P_\mathrm{s}}(0.95)$ is the fraction of users that successfully receive their desired signals (or the SINR is higher than the target threshold) in the first transmission attempt (i.e., $k = 1$) with a

probability higher than 0.95 (i.e., with reliability 0.95). $\bar{F}_{P_\mathrm{s}}(0.78)$ is the fraction of users that successfully receive their desired signals after the second transmission attempt (i.e., $k = 2$) with reliability 0.95. $\bar{F}_{P_\mathrm{s}}(0.63)$ is the fraction of users that successfully receive their desired signals after the third transmission attempt (i.e., $k = 3$) with reliability 0.95. $\bar{F}_{P_\mathrm{s}}(0.63)$ can also be interpreted as the fraction of users that successfully receive their desired signals in the first time slot with reliability 0.63, or the fraction of users that successfully receive their desired signals after the second time slot with reliability 0.86.

Chapter 3

Meta Distribution of SIR in Large-Scale Uplink and Downlink NOMA Networks

We develop an analytical framework to derive the meta distribution and moments of the conditional success probability (CSP), which is defined as success probability for a given realization of the transmitters, in large-scale co-channel uplink and downlink non-orthogonal multiple access (NOMA) networks with one NOMA cluster per cell. The moments of CSP translate to various network performance metrics such as the standard success or signal-to-interference ratio (SIR) coverage probability (which is the 1-st moment), the mean local delay (which is the -1-st moment in a static network setting), and the meta distribution (which is the complementary cumulative distribution function of the success or SIR coverage probability and can be approximated by using the 1-st and 2-nd moments). For the uplink NOMA network, to make the framework tractable, we propose two point process models for the spatial locations of the inter-cell interferers by utilizing the base station (BS)/user pair correlation function. We validate the proposed models by comparing the second moment measure of each model with that of the actual point process for the inter-cluster (or inter-cell) interferers obtained via simulations. For downlink NOMA, we derive closed-form solutions for the moments of the CSP, success (or coverage) probability, mean local delay, and meta distribution for the users.

3.1 Introduction

The next generations of wireless networks (such as 5G [56] or beyond 5G [B5G]) are expected to support billions of devices that are stimulated mainly from the diverse Internet-of-Things (IoT) applications (ranging from delay-tolerant machine-type communications (MTC) to delay sensitive mission-critical communications) in addition to the enhanced mobile broadband applications. As a result, acquiring ultra-reliable and low-latency communication (URLLC) is among one of the constitutional challenges for emerging massive wireless networks [57]. Traditionally, reliability can be achieved with efficient channel coding and retransmission schemes, e.g., hybrid automatic repeat request (HARQ). However, at the same time, massive device connectivity with strict latency requirements need to be achieved in URLLC systems. This necessitates efficient user access mechanisms that can potentially serve multiple devices in a specific time-frequency resource block while reducing their respective transmission delays [7].

Non-orthogonal multiple access (NOMA) has been recognized as a promising multi-user channel access technique that enables massive connectivity while reducing the transmission delay of the devices [7]. Contrary to traditional orthogonal multiple access (OMA), such as time division multiple access (TDMA), frequency division multiple access (FDMA), and code division multiple access (CDMA), the key idea of NOMA is to serve multiple users in the same channel simultaneously. The concurrent transmissions in NOMA reduce the waiting time of the devices while saving network resources. Of course, this can be achieved at the expense of additional interference and decoding complexity at the receivers. To mitigate interference, NOMA exploits successive interference cancellation (SIC) at the receivers [9, 8].

3.1.1 Background Work

Recently, performance analysis of NOMA-based wireless networks has attracted significant research interest. The existing studies contribute mainly toward understanding the **average performance** of users considering a single NOMA cell/cluster [8, 10, 11, 12, 13, 14]. For instance, the performance of a single-cell downlink NOMA system with randomly located users was first studied in [10]. In particular, the signal-to-interference-plus-noise ratio (SINR) outage probability and the ergodic capacity were derived for a user at rank m in terms of distance. In [11], the problem of user pairing was investigated considering fixed NOMA (F-NOMA) and cognitive radio inspired NOMA (CR-NOMA). A comparative performance analysis of uplink and downlink NOMA with selective two user pairing was conducted in [8]. Closed-form solutions for ergodic sum-rate and outage probability of a two-user NOMA cluster were presented in [12] considering a power back-off policy. The power back-off policy was applied to distinguish users in a NOMA cluster with nearly similar signal strengths (given that traditional uplink power control is in effect). The problem of user scheduling,

subcarrier allocation, and power control in uplink NOMA was investigated in [13, 14] with perfect SIC at the BS.

The aforementioned research studies ignore the impact of inter-cell interference which can significantly limit the performance of NOMA in massive wireless networks. Very recently, some of the research works have considered performance characterization of large-scale NOMA systems using stochastic geometry tools such as Poisson point process (PPP) and Poisson cluster process (PCP). The performance of uplink NOMA in terms of the rate coverage and average achievable rate was characterized first in [15] using PCP considering both perfect and imperfect SIC. For downlink NOMA, outage probability and average achievable rate of m-th rank user were derived in [16, 17] assuming that the BS locations follow a homogeneous PPP. The users are ranked based on their normalized channel gains defined as the channel gain including path-loss and small-scale fading normalized by the inter-cell interference. The analytical expressions are derived assuming that *the normalized channel gains of users located in a given NOMA cluster are independent and identically distributed (i.i.d.)*. However, since the inter-cell interferences received at the different users in the downlink are correlated, the normalized channel gains are also correlated, and therefore, the derived results are not precise. In [18], the performance of two-user downlink NOMA was investigated in a K-tier cellular network. The macro cell BSs use the massive multiple-input multiple-output (MIMO) technology and each small cell adopts user pairing to implement two-user NOMA transmission. In [19], for K-tier heterogeneous networks (HetNets) with biased nearest BS association, performance of downlink NOMA was investigated in terms of the coverage probability and throughput for non-cooperative and cooperative schemes. For Poisson cellular networks, [17] also studied the performance of uplink NOMA. To derive the analytical results, it was assumed that uplink inter-cell interferers form a homogeneous PPP which is inaccurate since there is no inter-cell interferer in the cell that the typical user lies in [20] (i.e., locations of the interferers are not independent). Moreover, unlike the downlink case, an exclusion zone around the serving BS (called the desired BS) of the typical user cannot be considered since an inter-cell interferer in the uplink can actually lie closer to the desired BS than the typical user. Therefore, the effective inter-cell interference field as observed from the desired BS can be modeled more accurately as an inhomogeneous PPP as demonstrated in [21, 22].

3.1.2 Motivation and Contributions

The current state-of-the-art mainly analyze the standard transmission success probability and ergodic capacity of users in NOMA-enabled cellular networks. Nevertheless, it is noteworthy that the standard transmission success probability is itself the mean of a random variable referred to as *conditional success probability (CSP)*, which is the success probability of a user considering a given realization of BSs [49]. When the

19

point processes describing the transmitter and receiver locations (referred to as transmitter point process and receiver point process) are ergodic, the standard success probability is the average of the CSPs of all users. Two networks can have the same standard (mean) success probability but distributions of the CSPs may be completely different. This is similar to the case where two different random variables have the same mean but different probability density functions (PDFs). To illustrate, let us consider the following example with two networks: network A and network B. Assume that, in network A, the CSP of half of the users is 0 while that of the other half is 1. In network B, the CSP is 0.5 for all the users. For these two networks, the standard (mean) success probability is the same and is equal to 0.5, while they behave completely different in terms of delay. In network A, half of the users have a successful transmission after their first attempt while the transmission delay is infinity for the other half of the users (i.e., the mean delay is infinity when CSP of users is 0 or close to 0). In network B, the number of transmission attempts until the first successful reception is almost surely finite for all the users, and for this network, the average number of transmission attempts for successful transmission is 2. Therefore, comparing two networks simply in terms of their average CSPs (or mean success probabilities) will not always be accurate since the CSP will not always be precisely characterized by its average value.

Along this line, [49] characterized the *meta distribution* which is the *complementary cumulative distribution function (CCDF) of the CSP*. This pioneering work was followed by various research studies for Poisson bipolar and cellular networks, device-to-device (D2D) networks, and millimeter-wave (mm-wave) D2D networks [49, 20, 51, 52, 53]. The meta distribution provides a more precise characterization of a typical transmission link than the standard success probability. Meta distribution enables us to answer questions such as "for a given target SIR threshold θ, what fraction of users are in coverage with reliability x?", whereas the standard success probability answers the question "for a given target SIR threshold θ, what fraction of users are in coverage?". Note that, CSP is a random variable and the meta distribution is the CCDF of this random variable while the standard success probability is its mean (average). Therefore, from the meta distribution we can derive the standard success probability, while the standard success probability does not reveal any information about the distribution of the CSP.

To this end, the contributions of this chapter can be summarized as follows:

- We derive the moments of CSP for uplink and downlink NOMA in Poisson cellular networks. This allows us to study the traditional success/coverage probability (which is the 1-st moment), the mean local delay (which is the -1-st moment), and the meta distribution (which is the CCDF of the conditional success probability and can be approximated by using the 1-st and 2-nd moments). Note that *mean local delay*, which is defined as the mean number of transmission attempts until the first

20

successful transmission [54], is a crucial performance metric for emerging URLLC systems.

- In uplink NOMA, the point process for the spatial locations of the interferers is a key for the derivation of the meta distribution and moments of CSP. Since the actual point process is unknown, we propose two models for this point process based on the pair correlation between interferers and the typical BS (which is at the origin). We demonstrate the accuracy of the proposed point processes by comparing the second moment measure of each process with that of the original process obtained via simulations. We show that the proposed point processes provide better approximations for derived results for low SIR threshold θ, and user locations closer to the BS.

- For downlink NOMA, we derive closed-form expressions for the moments of the CSP, success probability, mean local delay, and the meta distribution. We approximate the meta distribution by a beta distribution and demonstrate the accuracy of the approximation.

The rest of the chapter is structured as follows. In Section 3.2, we describe the system model and assumptions for uplink and downlink NOMA. In Section 3.3, for uplink NOMA, we propose two point processes to model the locations of the interferers, and derive the moments of the CSP and its meta distribution. In Section 3.4, for downlink NOMA, we derive closed-form solutions for the moments of the CSP and its meta distribution. Section 3.5 discusses numerical and simulation results followed by the conclusion in Section 3.6.

3.2 System Model and Assumptions

This section details the network model, channel model, and interference model along with assumptions for multi-user uplink and downlink NOMA systems.

3.2.1 Uplink NOMA

Network and Channel Model

We consider an uplink NOMA system where BSs are distributed according to a homogeneous PPP[1] Φ_{B} of intensity λ_{b}. Users are modeled as points of a homogeneous PPP of intensity λ_{u}. We assume a high user density regime, i.e., $\lambda_{\mathrm{u}} >> \lambda_{\mathrm{b}}$. Therefore, we can assume that there are at least N users located in each Voronoi cell. We consider random user selection, i.e., N users are randomly selected for NOMA transmission from the users located in the Voronoi cell[2]. Each user is connected to its nearest BS. The

---

[1]The motivation of modeling BS locations for real-world cellular networks with PPP was justified in [58].

[2]This can be viewed as the general case of the *user point process of type I* introduced in [22]. In [22], the user point process for $N = 1$ is studied which is the case in OMA. In this chapter, we consider $N \geq 1$.

network is interference-limited. The channel power between a user located at x and the typical BS located at the origin is given by $h_x \ell(x)$ where h_x represents the small-scale multi-path fading channel power following i.i.d. exponential distribution with unit mean and $\ell(x) = \|x\|^{-\alpha}$ represents the path-loss with exponent α, where $\alpha > 2$.

SIC

We assume that decoding order is based on the distance and consider perfect SIC, i.e., the BS perfectly decodes and cancels the first $m-1$ interference signals coming from rank 1 to $m-1$ users before decoding the signal of the m-th rank user[3]. The channel gains of different users are different[4] in the uplink, therefore each message signal experiences distinct channel gain. The conventional uplink power control, which is typically intended to equalize the received signal powers of users, may remove the channel distinctness and thus may not be feasible for uplink NOMA [15]. Therefore, we assume that all users transmit with the same power P.

Interference and SIR Model

To model the intra-cell interference with SIC, the typical BS needs to rank the received powers of various users. The ranking of users in terms of their distances from the serving BS is generally considered as a reasonable approximation of their respective ranked received signal powers[5] [59, 15]. This approximation provides tractability in the analysis. The intra-cell interference for the m-th rank user can therefore be modeled as:

$$I_{(m)}^{\text{intra}} = \sum_{i=m+1}^{N} P h_{x_{(i)}} \|x_{(i)}\|^{-\alpha}, \qquad m = 1, 2, ..., N, \tag{3.1}$$

where $h_{x_{(i)}}$ is the fading from the i-th rank user who is located at $x_{(i)}$ (in the Voronoi cell of the typical BS) to the typical BS. The inter-cell interference is given as follows:

$$I^{\text{inter}} = \sum_{x \in \Phi_{\text{I}}} P h_x \|x\|^{-\alpha}, \tag{3.2}$$

---

[3]In the next chapter, we introduce ordering methods and analyze the accuracy of distance-based ordering.

[4]The channel frequency/bandwidth is same for all users in NOMA; however, the channel gains experienced by the users on that specific frequency will be different due to their path-loss and fading.

[5]In Chapter 4, we further study this assumption. Removing the perfect SIC and distance-based ranking assumptions, in the next chapter, we derive the coverage probability for a more realistic setup.

where Φ_I is the point process describing the locations of the inter-cell interferers, which is unknown. Using (3.1) and (3.2), for the user at rank m, the SIR is given as follows:

$$\text{SIR}_{(m)} = \frac{P h_{x_{(m)}} \|x_{(m)}\|^{-\alpha}}{I_{(m)}^{\text{intra}} + I^{\text{inter}}}. \tag{3.3}$$

3.2.2 Downlink NOMA

Network and Channel Model

In downlink NOMA, the BSs apply superposition transmission and SIC is implemented at the receiver where each user decodes and cancels strong interferences before extracting the desired signal. A BS allocates more power to the weaker (located farther from the BS) users such that a user at rank m (ranked with respect to distance from the serving BS) can decode its signal in presence of interference from stronger users at rank 1 to $m-1$ [60, 8, 61, 62]. We consider N users in each Voronoi cell. The BSs are distributed according to a homogeneous PPP Φ_B of intensity λ_b and each BS can transmit with maximum power P_{BS}. The effect of thermal noise is neglected. The channel power gain between the BS located at x and the typical user located at the origin is given by $h_x \ell(x)$, and h_x for different users is modeled by i.i.d. exponential random variable with unit mean. $\ell(x) = \|x\|^{-\alpha}$ represents the power-law path-loss, in which α is the path-loss exponent. The power allocated to the i-th rank user is $P_i = \beta_i P_{\text{BS}}$, $\forall i = 1, 2, ..., N$. Also, we have $\beta_i \leq \beta_j$, $\forall i \leq j$ such that $\sum_{i=1}^{N} \beta_i = 1$ (or equivalently $\sum_{i=1}^{N} P_i = P_{\text{BS}}$).

SIC

We consider perfect SIC, i.e., user at rank m successfully removes the intra-cell interference of all users who are at higher ranks in terms of their distances.

In this chapter, as the first step towards performance analysis of NOMA in large-scale cellular networks, we assume perfect SIC and distance-based ordering to drive closed-form expressions. In the next chapter, we study SIC in more details. Specifically, we explain the difference between SIC in uplink and downlink, and analyze the accuracy of distance-based ranking in SIC.

Interference and SIR Model

The intra-cell interference at m-th rank user can be given as:

$$I_{(m)}^{\text{intra}} = \sum_{i=1}^{m-1} \beta_i P_{\text{BS}} h_0 \|x_0\|^{-\alpha}, \qquad m = 1, 2, ..., N, \tag{3.4}$$

where h_0 is the fading from the serving BS located at x_0 to the m-th rank user located at the origin. The inter-cell interference can be modeled as:

$$I_{(m)}^{\text{inter}} = \sum_{x \in \Phi_{\text{B}} \setminus \{x_0\}} P_{\text{BS}} h_x \|x\|^{-\alpha}, \qquad m = 1, 2, ..., N. \tag{3.5}$$

Hence, for the user at rank m, the SIR can be given as:

$$\text{SIR}_{(m)} = \frac{\beta_m P_{\text{BS}} h_0 \|x_0\|^{-\alpha}}{I_{(m)}^{\text{intra}} + I_{(m)}^{\text{inter}}}. \tag{3.6}$$

By Slivnyak's theorem, the point process for the inter-cell interferers is a PPP with intensity λ_{b} in $\mathbb{R}^2 \setminus b(o, \|x_0\|)$ where distribution of $\|x_0\|$ depends on the rank of the user. This is different from uplink where the inter-cell interference is received at a typical BS and is therefore same for all NOMA users in the typical Voronoi cell.

3.3   Uplink NOMA: Moments and Meta Distribution of the CSP

In this section, we derive the CSP and the moments and meta distribution of the CSP for an uplink NOMA network. For this, we first derive the distance distribution of the intra-cell interferers. Afterwards, we derive the approximate point process(es) of the inter-cell interferers, and then obtain the moments of CSP. The exact and approximate meta distributions can then be obtained by using (2.3) and (2.4), respectively, as described in Section 2.3. Let Φ_{U} denote the point process of the uplink NOMA users. Φ_{U} can be viewed as the superposition of two point processes, namely, intra-cell point process and inter-cell point process. The intra-cell point process consists of the uplink NOMA users that lie in the same cell as the typical uplink NOMA user (it also includes the typical uplink NOMA user). The inter-cell point process, denoted by Φ_{I}, consists of the inter-cell interferers.

For performance analysis of uplink NOMA, modeling the actual point process Φ_{I} for the inter-cell interferers is critical. In this section, we propose two point processes to approximate Φ_{I} since the exact process is not known. We will demonstrate the accuracy of the proposed point processes by comparing the second moment measure of each point process with the second moment measure of the actual (or original) point process Φ_{I} obtained by simulations. Since the typical BS is located at the o (origin) and we model Φ_{I} from the typical BS point of view, we are interested in the first and second moment measures for $b(o, r)$, where $b(o, r)$ denotes the ball of radius r centered at o.

3.3.1 Distance Distributions of the Intra-cell Interferers

The cell that contains the typical uplink NOMA user is the typical cell; therefore, the desired link distance between the typical uplink user and its serving BS is not Rayleigh distributed with mean $1/(2\sqrt{\lambda_{\rm b}})$ [22]. For the typical cell, the desired link distance between the typical uplink user and its serving BS can be approximated by a Rayleigh distribution with mean $1/\sqrt{5\lambda_{\rm b}}$ [20], which we follow in this chapter. Subsequently, for the typical uplink user, the probability density function (PDF) and cumulative distribution function (CDF) of the desired link distance are given as follows: [20]:

$$f_R(r) \approx (5/2)\lambda_{\rm b}\pi r e^{-(5/4)\lambda_{\rm b}\pi r^2}, \qquad F_R(r) \approx 1 - e^{-(5/4)\lambda_{\rm b}\pi r^2}. \tag{3.7}$$

Using the above distributions and order statistics [63, 64], the distribution of the distance of the user at rank m from its serving BS can be derived as follows:

$$f_{R_m}(r) = \frac{5\lambda_{\rm b}\pi r \left(1 - e^{-(5/4)\lambda_{\rm b}\pi r^2}\right)^{m-1} \left(e^{-(5/4)\lambda_{\rm b}\pi r^2}\right)^{N-m+1}}{2B(N-m+1, m)}, \qquad r \geq 0, \tag{3.8}$$

where $B(\cdot, \cdot)$ is the Beta function. Conditioned on the distance of the user at rank m $(R_m = r_m)$, it was shown in [65, 15] that the distances of users at lower or higher ranks than m-th rank user to the typical BS are i.i.d. and their PDFs can be characterized, respectively, as follows:

$$f_{R_{\rm in}}(r \mid R_m = r_m) = \frac{f_R(r)}{F_R(r_m)}, \qquad\qquad r \leq r_m, \quad i = 1, ..., m-1, \tag{3.9}$$

$$f_{R_{\rm out}}(r \mid R_m = r_m) = \frac{f_R(r)}{1 - F_R(r_m)}, \qquad\qquad r \geq r_m, \quad i = m+1, ..., N. \tag{3.10}$$

3.3.2 BS/user Pair Correlation Function

In this section, we proceed by obtaining BS/user pair correlation function which is the pair correlation function of $\Phi_{\rm I}$ with respect to the origin (the location of the typical BS), through simulations, and then model the locations of the inter-cell interferers by proposing two point processes with the same BS/user pair correlation function.

The pair correlation function measures the correlation (dependency) between the points of a point process. Similar to [22], we study the cross-correlation between the inter-cell interfering users' point process and the typical BS. Therefore, we use the term "BS/user pair correlation function"[6]. For a motion invariant point

---

[6]BS/inter-cell user pair correlation function is more accurate. However, to be consistent with [22], we also use the term "BS/user pair correlation function".

process Φ with intensity λ, the pair correlation function is defined as [41]:

$$g(r) = \frac{1}{2\pi r} \frac{\mathrm{d}}{\mathrm{d}r} \left(\frac{1}{\lambda} \mathbb{E}^{0!}[\Phi(b(o,r))] \right),$$

where $\mathbb{E}^{0!}$ is the reduced palm expectation given that a point exists at the origin. In the following, we will use the above definition and replace the reduced palm expectation with $\mathbb{E}^{0}[\Phi_{\mathrm{I}}(b(o,r))]$ which is the mean number of inter-cell interferers that are located within distance r from the typical BS which is located at the origin. Therefore, this term captures the cross-correlation between the typical BS and Φ_{I}.

Definition 3.1 (BS/user Pair Correlation Function [22]). *For the motion invariant BS and user point processes, the BS/user pair correlation function $g_{\lambda_{\mathrm{b}}}(r)$ is defined as follows:*

$$g_{\lambda_{\mathrm{b}}}(r) \triangleq \frac{1}{2\pi r} \frac{\mathrm{d}}{\mathrm{d}r} K(r) = \frac{1}{2\pi r} \frac{\mathrm{d}}{\mathrm{d}r} \left(\frac{1}{N\lambda_{\mathrm{b}}} \mathbb{E}^{0}[\Phi_{\mathrm{I}}(b(o,r))] \right), \tag{3.11}$$

where $\mathbb{E}^{0}$ is the Palm expectation (given that the typical BS is at the origin), and λ_{b} is the intensity of the BS point process. When Φ_{I} is scale-invariant, $g_{\lambda_{\mathrm{b}}}(r) = g_1(\sqrt{\lambda_{\mathrm{b}}}r)$.

Note that the BS/user pair correlation function $g_{\lambda_{\mathrm{b}}}(r)$ is useful in approximating the inter-cell interfering users' point process by a PPP of intensity function $\lambda_{\mathrm{b}} g_{\lambda_{\mathrm{b}}}(r)$ [22]. Specifically, [22] studied the point process of uplink interferers for $N = 1$ (i.e., for orthogonal multiple access [OMA]), and through numerical fitting, the best exponential fit for $N = 1$ was obtained as follows:

$$g_1(r) = 1 - e^{-(12/5)\pi r^2}. \tag{3.12}$$

Note that any other point process with the same intensity function ($\lambda_{\mathrm{b}} g_{\lambda_{\mathrm{b}}}(r)$) can also be used to approximate the inter-cell interferers' point process.

Along the same lines, we also obtain $g_1(r)$ through simulations. For $N = 2$ and $N = 5$, $g_1(r)$ is illustrated in Fig. 3.1 and we observe that it does not vary for any value of N. The reason is that the average number of inter-cell interferers within distance r from the typical BS $\mathbb{E}^{0}[\Phi_{\mathrm{I}}(b(o,r))]$ for clusters of N users in NOMA is N times higher than OMA. Therefore, (3.11) does not change with respect to N. We also compare the simulation results with the best exponential fit for $N = 1$. Using the invariance property of $g_1(r)$ with respect to N, the scale-invariance property of the model, and the results in [22], we approximate the inter-cell interferers Φ_{I} by a PPP $\bar{\bar{\Phi}}_{\mathrm{I}}$ with intensity $N\lambda_{\mathrm{b}} g_{\lambda_{\mathrm{b}}}(r)$. NOMA users in each NOMA cluster must be served by the same BS, i.e, they must be in the same Voronoi cell. Therefore, distances among the users is limited by the size of the Voronoi cell and the users' locations are correlated. Since, in a PPP, the location
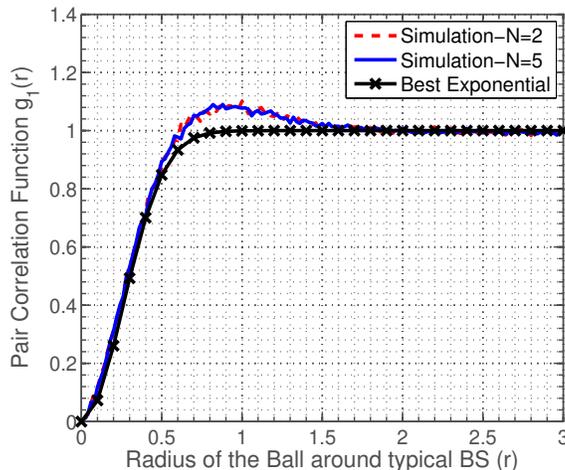
Figure 3.1: BS/user pair correlation function $g_1(r)$ for different N and validation with the best exponential fit in (3.12) for $N=1$.

of each point is independent of the other points [41], this correlation cannot be modeled accurately with $\bar{\bar{\Phi}}_{\mathrm{I}}$. To address this issue, we also propose a cluster process $\bar{\Phi}_{\mathrm{I}}$ to approximate Φ_{I}. In the following, first we define the intensity measure and then we describe the two proposed models along with their validation and comparative analysis.

Definition 3.2 (Intensity Measure [41])**.** *For any point process Φ, the intensity measure (first moment measure) $\Lambda(B)$ is the mean number of points in B, i.e., $\Lambda(B) = \mathbb{E}[\Phi(B)]$, $\quad \forall B \subset \mathbb{R}^2$. If Φ has an intensity function $\lambda(x)$, then $\Lambda(B) = \int_B \lambda(x)\mathrm{d}x$.*

3.3.3 Point Process Models for Inter-cell Interferers

Model 1

To model the interferers, we consider a PCP $\bar{\Phi}_{\mathrm{I}}$[7], where the parents form an inhomogeneous PPP $\bar{\Phi}_{\mathrm{P}}$ with intensity function $\bar{\lambda}_{\mathrm{p}}(x) = \lambda_{\mathrm{b}}\left(1 - e^{-(12/5)\lambda_{\mathrm{b}}\pi\|x\|^2}\right)$. In each cluster, N offspring points are located in the same location as the parent, i.e., for a parent at x, N offsprings are i.i.d. with PDF $f(y) = \delta(y - x)$, where $x, y \in \mathbb{R}^2$. This model can also be viewed as a non-simple inhomogeneous PPP [41]. As mentioned earlier, other cluster processes that have the same BS/user pair correlation function can also be used to model the inter-cell interferers, but $\bar{\Phi}_{\mathrm{I}}$ is more tractable. Note that $\bar{\Phi}_{\mathrm{I}}$ models the locations of the inter-cell interferers. **Model 1** implies that, although inter-cell interferers are located in different locations in each Voronoi cell, for large-scale analysis of uplink NOMA, it can be assumed that, from the point of view of the typical BS, they are co-located.

---

[7]If the parents of a cluster process are the points of a Poisson process, the resulting process is a Poisson cluster process [41].

Using **Model 1**, the mean number of inter-cell interferers within the distance r from the typical BS (first moment of $\bar{\Phi}_{\mathrm{I}}(b(o,r))$) can be derived as follows:

$$\bar{\Lambda}(b(o,r)) = \mathbb{E}\left[\bar{\Phi}_{\mathrm{I}}(b(o,r))\right] \stackrel{(a)}{=} N\mathbb{E}\left[\bar{\Phi}_{\mathrm{P}}(b(o,r))\right] = N\bar{\Lambda}_{\mathrm{p}}(b(o,r))$$

$$= N \int_{b(o,r)} \bar{\lambda}_{\mathrm{p}}(x)\mathrm{d}x = N\lambda_{\mathrm{b}}\left[\pi r^2 - \frac{5}{12\lambda_{\mathrm{b}}}\left(1 - e^{-(12/5)\lambda_{\mathrm{b}}\pi r^2}\right)\right], \tag{3.13}$$

where $\bar{\Lambda}$ and $\bar{\Lambda}_{\mathrm{p}}$ are the intensity measures of $\bar{\Phi}_{\mathrm{I}}$ and $\bar{\Phi}_{\mathrm{P}}$, respectively. Step (a) follows from $\bar{\Phi}_{\mathrm{I}}(b(o,r)) = N\bar{\Phi}_{\mathrm{P}}(b(o,r))$. The second moment measure of $\bar{\Phi}_{\mathrm{I}}b(o,r))$ is derived as follows:

$$\mathbb{E}\left[\bar{\Phi}_{\mathrm{I}}^2(b(o,r))\right] = N^2\mathbb{E}\left[\bar{\Phi}_{\mathrm{P}}^2(b(o,r))\right] \stackrel{(a)}{=} N^2\sum_{k=0}^{\infty} k^2 \frac{\bar{\Lambda}_{\mathrm{p}}(b(o,r))^k}{k!}e^{-\bar{\Lambda}_{\mathrm{p}}(b(o,r))}$$

$$\stackrel{(b)}{=} N^2\left[\bar{\Lambda}_{\mathrm{p}}(b(o,r)) + \bar{\Lambda}_{\mathrm{p}}(b(o,r))^2\right] = \bar{\Lambda}(b(o,r))\left[N + \bar{\Lambda}(b(o,r))\right], \tag{3.14}$$

where (a) follows since $\bar{\Phi}_{\mathrm{P}}(b(o,r))$ is a Poisson random variable with mean $\bar{\Lambda}_{\mathrm{p}}(b(o,r))$, and (b) is obtained from mean and variance of the Poisson distribution, where $\bar{\Lambda}(b(o,r))$ is given in (3.13). (3.14) can also be derived using the second factorial moment measure of PPPs.

Model 2

We approximate Φ_{I} with an inhomogeneous PPP $\bar{\bar{\Phi}}_{\mathrm{I}}$ with intensity function $\bar{\bar{\lambda}}(x) = N\lambda_{\mathrm{b}}\left(1 - e^{-(12/5)\lambda_{\mathrm{b}}\pi\|x\|^2}\right)$. The mean number of inter-cell interferers within the distance r from the typical BS (first moment of $\bar{\bar{\Phi}}_{\mathrm{I}}(b(o,r))$) is as follows:

$$\bar{\bar{\Lambda}}(b(o,r)) = \mathbb{E}\left[\bar{\bar{\Phi}}_{\mathrm{I}}(b(o,r))\right] = \int_{b(o,r)} \bar{\bar{\lambda}}(x)\mathrm{d}(x) = \bar{\Lambda}(b(o,r)),$$

where $\bar{\bar{\Lambda}}$ denotes the intensity measure of $\bar{\bar{\Phi}}_{\mathrm{I}}$. The second moment of $\bar{\bar{\Phi}}_{\mathrm{I}}(b(o,r))$ is given by

$$\mathbb{E}\left[\bar{\bar{\Phi}}_{\mathrm{I}}^2(b(o,r))\right] \stackrel{(a)}{=} \bar{\bar{\Lambda}}(b(o,r))\left[\bar{\bar{\Lambda}}(b(o,r)) + 1\right], \tag{3.15}$$

where (a) follows since $\bar{\bar{\Phi}}_{\mathrm{I}}(b(o,r)$ is Poisson variable with mean $\bar{\bar{\Lambda}}(b(o,r))$. Note that the point process introduced in [22] is a special case ($N = 1$) of the proposed **Models 1** and **2**.

According to the simulation results for success probability and the meta distribution in Section 3.5, analytical results that are obtained from **Model 1** provide upper bounds while those that are derived from **Model 2** provide lower bounds. Moreover, **Model 2** provides a better approximation for the nearest user ($m = 1$), while **Model 1** provides a better approximation for the furthest user $m = N$.
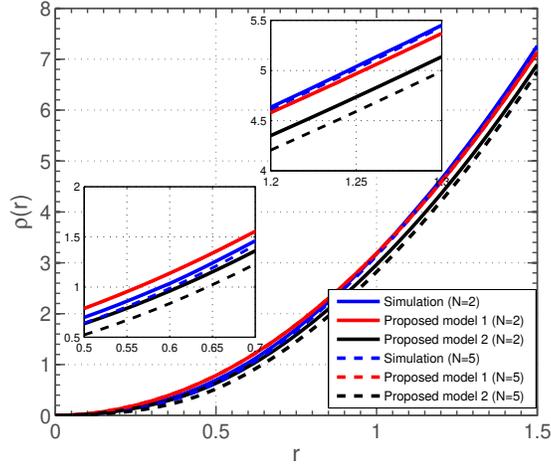
Figure 3.2: Comparing the second moment measure of Φ_I and the proposed models (derived in (3.14) and (3.15)) for $\lambda_b = 1$.

3.3.4 Model Validation

To compare the second moment of $\Phi_I(b(o,r))$ with the proposed models, we define $\rho(r) \triangleq \frac{1}{N\lambda_b}\sqrt{\mathbb{E}\left[\Phi_I^2(b(o,r))\right]}$.[8] In Fig. 3.2, $\rho(r)$ for the original inter-cell interferers' point process Φ_I is obtained via simulations and a comparison is performed with the proposed models. We observe that, $\rho(r)$ for the proposed models are close to the $\rho(r)$ of Φ_I. Moreover, based on Fig. 3.2, **Model 1** provides a better approximation for larger values of r.

3.3.5 Moments and Meta Distribution of the CSP

The moments of the CSP for uplink NOMA users can be derived as follows.

Theorem 3.1 (Moments of the CSP for Uplink NOMA). *In uplink NOMA, the b-th moment of the CSP, $b \in \mathbb{C}$, for the m-th rank user can be derived as follows:*

$$M_{b,(m)} = \int_0^\infty \left[(5/2)\lambda_b\pi r^2 e^{(5/4)\lambda_b\pi r^2}\mu_b\left((5/4)\lambda_b\pi r^2,\theta\right)\right]^{N-m} \mathbb{E}\left[\prod_{x\in\Phi_I}\left(\frac{1}{1+\theta r^\alpha\|x\|^{-\alpha}}\right)^b\right]f_{R_m}(r)\mathrm{d}r, \tag{3.16}$$

*where $f_{R_m}(r)$ is given in (3.8) and $\mu_b(x,z) = \int_0^1 \frac{t^{-3}e^{-xt^{-2}}}{(1+zt^\alpha)^b}\mathrm{d}t$. The expectation in (3.16), which is conditioned on the serving distance r, can be approximated by using the proposed **Model 1** and **Model 2**, respectively,*

---

[8]We consider the square root of the normalized second moment since it illustrates the difference between the models better.

as follows:

$$\mathbb{E}\left[\prod_{x\in\Phi_{\mathrm{I}}}\left(\frac{1}{1+\theta r^{\alpha}\|x\|^{-\alpha}}\right)^{b}\right] \approx \exp\left\{-2\pi\lambda_{\mathrm{b}}\int_{0}^{\infty}\left[1-\left(\frac{1}{1+\theta r^{\alpha}x^{-\alpha}}\right)^{Nb}\right]\left(1-e^{-(12/5)\lambda_{\mathrm{b}}\pi x^{2}}\right)x\mathrm{d}x\right\}, \quad (3.17)$$

$$\mathbb{E}\left[\prod_{x\in\Phi_{\mathrm{I}}}\left(\frac{1}{1+\theta r^{\alpha}\|x\|^{-\alpha}}\right)^{b}\right] \approx \exp\left\{-2\pi N\lambda_{\mathrm{b}}\int_{0}^{\infty}\left[1-\left(\frac{1}{1+\theta r^{\alpha}x^{-\alpha}}\right)^{b}\right]\left(1-e^{-(12/5)\lambda_{\mathrm{b}}\pi x^{2}}\right)x\mathrm{d}x\right\}. \quad (3.18)$$

Proof. See **Appendix A.1**. □

As shown below, the proposed point processes provide better approximations for standard transmission success probability ($b=1$) when the SIR threshold is low.

Corollary 3.1. *For $b=1$ (standard success probability), the proposed point process models provide better approximations when $\theta \to 0$ compared to large values of θ.*

Proof. For $b = 1$, we have $\mathbb{E}\left[\prod_{x\in\Phi_{\mathrm{I}}}\frac{1}{1+\theta r^{\alpha}\|x\|^{-\alpha}}\right] = L_{I^{\mathrm{inter}}}(\theta r^{\alpha}/P)$, where I^{inter} is given in (3.2) and $L_{I^{\mathrm{inter}}}(s) = \mathbb{E}\left[e^{-sI^{\mathrm{inter}}}\right]$ is the Laplace transform of the inter-cell interference. When $s \to 0$, we have

$$L_{I^{\mathrm{inter}}}(s) = \mathbb{E}\left[e^{-sI^{\mathrm{inter}}}\right] \sim 1 - \mathbb{E}\left[sI^{\mathrm{inter}}\right] \stackrel{(a)}{=} 1 - \mathbb{E}\left[s\sum_{x\in\Phi_{\mathrm{I}}}Ph_{x}\|x\|^{-\alpha}\right] \stackrel{(b)}{=} 1 - sP\mathbb{E}\left[\sum_{x\in\Phi_{\mathrm{I}}}\|x\|^{-\alpha}\right],$$

where (a) follows from (3.2), and (b) follows since the fading coefficients h_{x} are i.i.d. with unit mean.

According to the Campbell's theorem, approximating Φ_{I} with point processes that have the same BS/user pair correlation function (which can also be interpreted as the intensity measure with respect to the origin) for any $f : \mathbb{R}^{2} \mapsto \mathbb{R}$ yields

$$\mathbb{E}\left[\sum_{x\in\Phi_{\mathrm{I}}}f(x)\right] \equiv \mathbb{E}\left[\sum_{x\in\bar{\Phi}_{\mathrm{I}}}f(x)\right] \equiv \mathbb{E}\left[\sum_{x\in\bar{\bar{\Phi}}_{\mathrm{I}}}f(x)\right].$$

Therefore, proposed models provide better approximations for the first moment M_{1} when $\theta \to 0$. □

Similarly, we can prove that for $m = 1$ the approximations are better compared to other values of m, because the probabilities of small values of r are higher for $m = 1$. Moreover, after applying changes of variables and some simplifications in (3.16), we can show that the moments of the CSP in uplink NOMA are independent of the BS intensity λ_{b}.

Corollary 3.2. *For $b \in \mathbb{R}$, $M_{b,(m)}$ of **Model 2** is a lower bound for $M_{b,(m)}$ of **Model 1**.*

Proof. Using the identity $1 - y^N \equiv (1-y)\left(1 + y + y^2 + ... + y^{N-1}\right)$, for $0 \leq y$, we have, $1 - y^N \leq N(1-y)$. Then **Corollary 3.2** is obtained by setting $y = \left(\frac{1}{1+\theta r^\alpha x^{-\alpha}}\right)^b$ in (3.17) and (3.18). $\qquad\qquad\square$

The exact and approximate meta distributions of CSP can be obtained by using (2.3) and (2.4), respectively, as described in Section 2.3.

Remark: Perfect SIC offers an optimistic view of the system performance especially for scenarios where the users are located very close to each other. However, these scenarios are less likely to occur when the number of users in a NOMA cluster is small, which would be usually the case in a practice. The performance gains of NOMA over OMA are generally achievable for $N = 2$ or $N = 3$ and in such scenarios interference cancellation is therefore likely to be achieved. Thus, achieving near perfect SIC is not completely unrealistic and has been considered in most of the existing state-of-the-art work [66, 67, 68, 69, 70]. Nevertheless, we can characterize the system performance with the worst-case SIC, where a given user will be able to decode its signal only when all the strong interferences are decoded perfectly. This will provide a pessimistic view (i.e., lower bound) of the system performance.[9]

In uplink NOMA with the worst-case SIC, decoding of the desired signal of the user at rank m is always unsuccessful whenever the decoding of the $m - 1$ closer users (strong interferences) are unsuccessful. Therefore, the success probability of the m-th rank user is

$$p_{(m)}^{\text{worst-case}} = \prod_{i=1}^{m} M_{1,(i)},$$

where $M_{1,(i)}$ is given in **Theorem 3.1**. The corresponding expression for the downlink case will be provided in the next section.

3.4 Downlink NOMA: Moments and Meta Distribution of the CSP

In this section, we derive the CSP, and the moments and meta distribution of the CSP in a downlink NOMA network. For this, we first derive the distance distribution of the desired link and then derive the moments of CSP as well as the meta distribution.

---

[9]Analysis of the worst-case SIC for both uplink and downlink with the impact of interference correlation is postponed to the next chapter.

3.4.1 Distribution of the Desired Link Distance

Since each user connects to its nearest BS, the serving link distance distribution is given by the Rayleigh distribution as follows [71]:

$$f_R(r) = 2\lambda_b \pi r e^{-\lambda_b \pi r^2}, \quad F_R(r) = 1 - e^{-\lambda_b \pi r^2}. \tag{3.19}$$

Using the above equations and order statistics [63, 64], the distribution of the distance of a user at rank m from its serving BS is as follows:

$$f_{R_m}(r) = \frac{2\lambda_b \pi r \left(1 - e^{-\lambda_b \pi r^2}\right)^{m-1} \left(e^{-\lambda_b \pi r^2}\right)^{N-m+1}}{B(N-m+1, m)}, \quad r \geq 0. \tag{3.20}$$

3.4.2 Moments and Meta Distribution of the CSP

The b-th moment of the conditional success probability $M_{b,(m)}$, $b \in \mathbb{C}$, for an m-th rank downlink NOMA user is derived in the following. Based on these moments, we can derive the mean success probability, the meta distribution, and the mean local delay.

Theorem 3.2 (Moments of the CSP for Downlink NOMA). *For a user at rank m, the b-th moment of the conditional success probability $M_{b,(m)}$ is*

$$M_{b,(m)} = \begin{cases} \frac{B(A_{b,m}+N-m+1,m)}{B(N-m+1,m)}, & \theta < \beta_m / \sum\limits_{i-1}^{m-1} \beta_i \\ 0, & \theta > \beta_m / \sum\limits_{i-1}^{m-1} \beta_i \quad \& \quad \Re(b) > 0 \\ \infty, & \theta > \beta_m / \sum\limits_{i-1}^{m-1} \beta_i \quad \& \quad \Re(b) < 0 \end{cases} \tag{3.21}$$

where $A_{b,m} = \sum_{k=1}^{\infty} \binom{b}{k}(-1)^{k+1} c_m^k \frac{\delta}{k-\delta} {}_2F_1(k, k-\delta; k-\delta+1; -c_m)$, $c_m = (\frac{\beta_m}{\theta} - \sum\limits_{i-1}^{m-1} \beta_i)^{-1}$, $\delta = 2/\alpha$, ${}_2F_1$ is the Gauss hypergeometric function, and $\Re(b)$ gives the real part of b.

Proof. See **Appendix A.2**. $\square$

Note that, similar to the uplink, moments of the CSP in downlink are independent of the BS intensity λ_b. The condition $\theta < \beta_m / \sum\limits_{i-1}^{m-1} \beta_i$ (or equivalently, $c_m > 0$) implies that the received SIR at the m-th rank user is greater than the required SIR θ in the absence of inter-cell interference. Moreover, when $N = 1$, which is the case in orthogonal multiple access, **Theorem 3.2** reverts back to the known results for downlink Poisson cellular networks [49].

With the worst-case SIC, decoding of the desired signal of the user at rank m, is always unsuccessful whenever the decoding of the $N - m$ farther users (strong interferences) are unsuccessful. Therefore, the success probability of the m-th rank user is

$$p_{(m)}^{\text{worst-case}} = \prod_{i=m}^{N} \frac{B(A_{1,i} + N - m + 1, m)}{B(N - m + 1, m)},$$

where $A_{1,i}$ is given in **Theorem 3.2**. Note that, in downlink, each user performs decoding separately while in uplink, detection of signals is performed at the BS.

In the following, a simplified closed-form expression for negative moments $M_{-w,(m)}$, $w \in \mathbb{R}^+$, is provided. The expression is useful in evaluating the mean local delay of an m-th rank user in closed-form by setting $w = 1$.

Corollary 3.3. *When $b = -w$, $w \in \mathbb{R}^+$, and $c_m > 0$*

$$M_{-w,(m)} = \begin{cases} \frac{B(N-m-D_{w,m}+1,m)}{B(N-m+1,m)}, & D_{w,m} < N - m + 1 \\ \\ \infty, & \text{otherwise} \end{cases} \tag{3.22}$$

*where $D_{w,m} = \sum_{k=1}^{\infty} \binom{w}{k} c_m^k \frac{\delta}{k-\delta}$. When $c_m < 0$, from **Theorem 3.2**, we have $M_{-w,(m)} = \infty$.*

Proof. From the proof of **Theorem 3.2**, we have

$$M_{-w,(m)} = \mathbb{E}_{R_m} \left[\exp \left\{ -2\pi\lambda_{\text{b}} \int_{R_m}^{\infty} \left[1 - \left(\frac{1}{1 + c_m R_m^\alpha r^{-\alpha}} \right)^{-w} \right] r \mathrm{d}r \right\} \right]$$

$$= \int_0^{\infty} \exp \left\{ \pi\lambda_{\text{b}} r^2 \sum_{k=1}^{\infty} \binom{w}{k} c_m^k \frac{\delta}{k - \delta} \right\} f_{R_m}(r) \mathrm{d}r. \tag{3.23}$$

Finally, **Corollary 3.3** is obtained by substituting (3.20) in (3.23) and setting $D_{w,m} = \sum_{k=1}^{\infty} \binom{w}{k} c_m^k \frac{\delta}{k-\delta}$. $\qquad \square$

Note that setting $b = jt$, $j = \sqrt{-1}$ and $t \in \mathbb{R}^+$, the exact meta distribution of the m-th rank user is derived by substituting $M_{jt,(m)}$ from **Theorem 3.2** in (2.3).

Since the exact meta distribution is complicated and does not provide any direct insights, the corresponding beta approximation can be obtained. To derive the beta approximation, we need the first and second moments of $P_{\text{s},(m)}$. Standard (mean) success probability which is the first moment of $P_{\text{s},(m)}$ can be easily obtained by setting $b = 1$ in **Theorem 3.2**, i.e., $M_{1,(m)} = B(A_{1,m} + N - m + 1, m)/B(N - m + 1, m)$, where $A_{1,m} = c_m \frac{\delta}{1-\delta} \, _2F_1(1, 1 - \delta; 2 - \delta; -c_m)$. Similarly, we can derive the second moment $M_{2,(m)}$ by setting $b = 2$. The beta approximation is obtained by substituting $M_{1,(m)}$ and $M_{2,(m)}$ in (2.4) as described in Section 2.3.

(a) Standard success probability for $N = 2$.   (b) Meta distribution for $N = 3$.
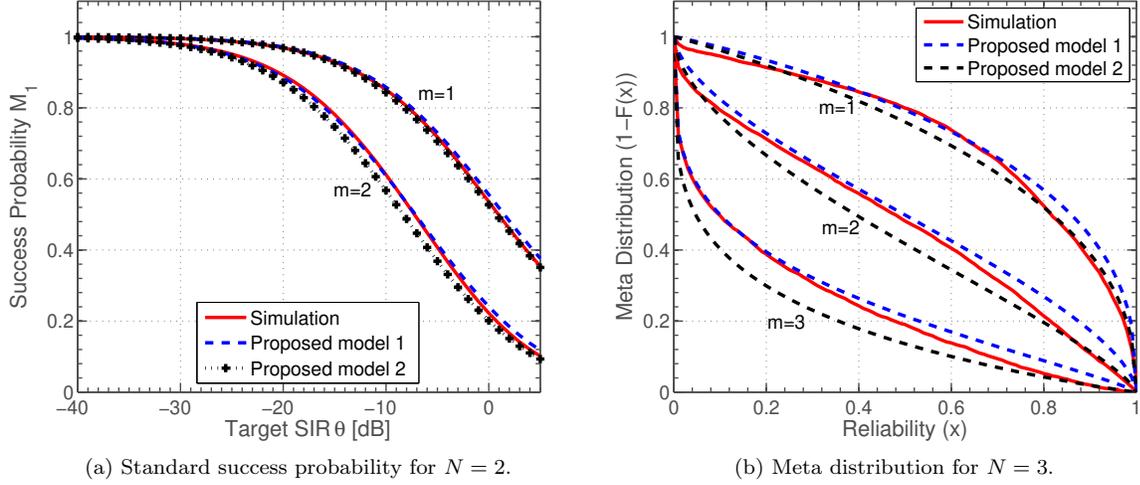
Figure 3.3: First moment of the CSP (i.e., standard success probability) and meta distribution for uplink NOMA with $\lambda_b = 0.001$ and $\alpha = 4$.

Corollary 3.4 (Mean Local Delay of User at Rank m). *For the m-th rank user, when $c_m > 0$ and $D_{1,m} = c_m \frac{\delta}{1-\delta} < N - m + 1$, the mean local delay is finite and is given by*

$$M_{-1,(m)} = \frac{B(N - m - D_{1,m} + 1, m)}{B(N - m + 1, m)}. \tag{3.24}$$

When $c_m < 0$ or $D_{1,m} = c_m \frac{\delta}{1-\delta} > N - m + 1$, the mean local delay is infinite.

3.5 Numerical Results and Discussions

In this section, we present numerical and simulation results to validate the accuracy of the derived expressions. Specifically, for uplink NOMA, we validate and compare the analytical results of **Theorem 3.1** considering the two proposed models for the inter-cell interferers' point process. A comparison is also provided with the traditional OMA scheme. For both uplink and downlink NOMA, we validate the accuracy of the beta approximation for the meta distribution using the results in **Theorem 3.1** and **Theorem 3.2** and show the distribution of the CSP for different users in a NOMA cluster.

3.5.1 Uplink NOMA

Validation of Model 1 and Model 2 and Meta Distribution of CSP

To demonstrate the accuracy of the proposed inter-cell interferers' point process models, in Fig. 3.3(a), we plot the first moment of the CSP, which is the standard success probability, of a user at rank m. Simulation
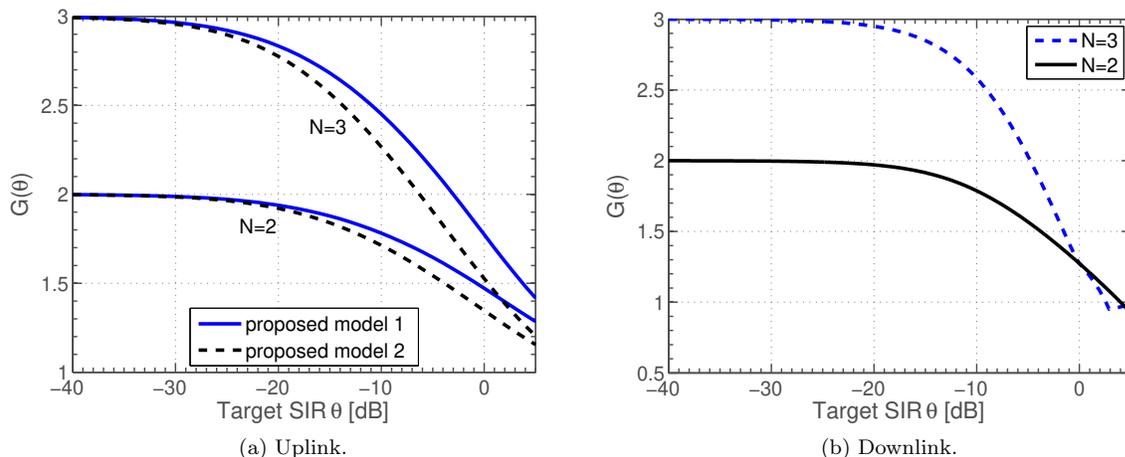
Figure 3.4: $G(\theta)$ for uplink and downlink NOMA. (a) Uplink NOMA with $\lambda_{\mathrm{b}} = 0.0005$ and $\alpha = 4$. (b) Downlink NOMA with $\lambda_{\mathrm{b}} = 0.001$ and $\alpha = 4$. For $N = 2$, $\beta_1 = 0.15$ and $\beta_2 = 0.85$. For $N = 3$, $\beta_1 = 0.17$, $\beta_2 = 0.33$, and $\beta_3 = 0.5$.

results and the analytical results derived in **Theorem 3.1** are compared for $\lambda_{\mathrm{b}} = 0.001$, $N = 2$, and $\alpha = 4$. According to Fig. 3.3, **Model 2** provides a better approximation for $m = 1$, while **Model 1** provides a better approximation for $m = N$. In general, **Model 1** outperforms in a wide range of scenarios.

For the same BS intensity and path-loss exponent with three users in each NOMA cluster, the exact meta distribution of the CSP (obtained via simulations) and its beta approximation (with two approximate inter-cell interferer point processes) are shown in Fig. 3.3(b). The beta distribution as well as the proposed point process models provide a good approximation for the exact meta distribution.

NOMA vs. OMA

To compare N-user NOMA with OMA, we define the gain G as

$$G(\theta) \triangleq \frac{\sum_{m=1}^{N} M_{1,(m)}(\theta)}{M_1^{\mathrm{OMA}}(\theta)}, \tag{3.25}$$

where M_1^{OMA} considers no channel inversion power control and is obtained by setting $N = m = 1$ in **Theorem 3.1**. For a given amount of radio bandwidth, when BS and user point processes are ergodic, $G(\theta)$ can be interpreted as the ratio of the density of users served in NOMA to the density of users served in OMA. For instance, according to Fig. 3.4(a), when $N = 3$, $G(-10\text{ dB}) \approx 2.3$, which means, with NOMA, the number of users served in a unit area is 2.3 times that with OMA. In Fig. 3.4(a), the gain of uplink NOMA $G(\theta)$ decays rapidly with increasing θ and the rate of decay is much higher for large number of users N.

(a) $\beta_1 = 0.17$, $\beta_2 = 0.33$, and $\beta_3 = 0.5$.　　　　(b) $\beta_1 = 0.08$, $\beta_2 = 0.25$, and $\beta_3 = 0.67$.
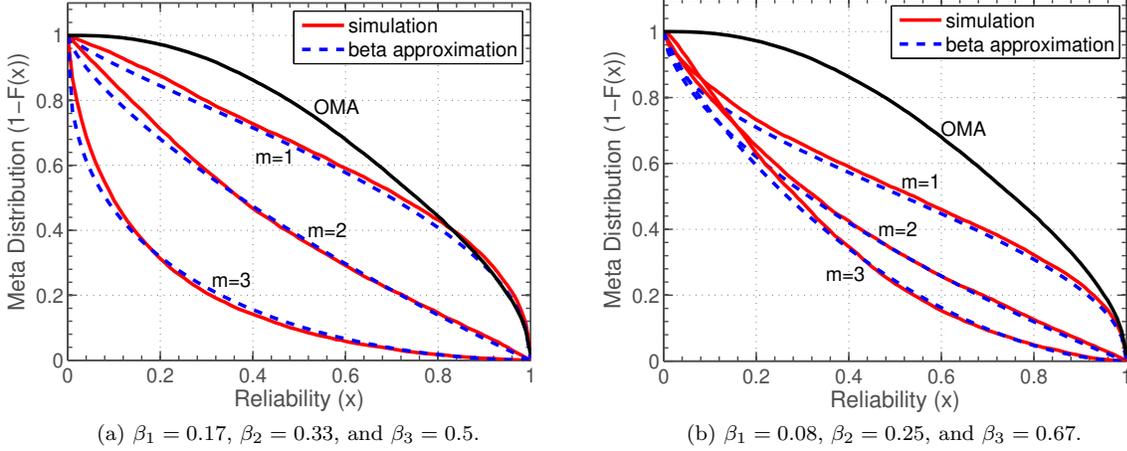
Figure 3.5: The exact meta distribution and its beta approximation for m-th rank user when $\lambda_{\mathrm{b}} = 0.001$, $N = 3$, $\alpha = 4$, and $\theta = -3\,\mathrm{dB}$.

3.5.2　Downlink NOMA

NOMA vs. OMA

In Fig. 3.4(b), $G(\theta)$ is evaluated for downlink. Similar to the uplink, the gain of downlink NOMA $G(\theta)$ decays rapidly with increasing θ and the rate of decay is much higher for large number of users N. However, for large values of θ, $G(\theta)$ increases since the effect of link distance is dominant and the average link distance of a typical user in OMA is $1/(2\sqrt{\lambda_{\mathrm{b}}})$, while in NOMA, the average link distance of the 1-st rank user is $1/(2\sqrt{N\lambda_{\mathrm{b}}})$.

Validation of Meta Distribution of CSP

In Fig. 3.5, we show that the meta distribution for the CSP can be approximated by the beta distribution with shape parameters $M_1\beta/(1 - M_1)$ and β. We consider three users in each NOMA cell. In this scenario, the meta distribution of the m-th rank user, $m = 1, 2, 3$, and its beta approximation are shown in Fig. 3.5 for two different power allocations. It can be seen that the beta distribution provides a good approximation for the meta distribution. In Fig. 3.5(a), we note that about 58% of the 1-st rank users, 30% of 2-nd rank users, and 7% of 3-rd rank users have success probabilities greater than 0.6. Therefore, success probabilities of 32% of users are greater than 0.6. With OMA, for 68% of users, success probabilities are greater than 0.6. This means that, with NOMA, the density of users served with the same amount of radio spectrum is $32 \times 3/68 \approx 1.4$ times that with OMA, when the target reliability is 0.6.

Using (2.6), we can also study the distribution of the local delay from Fig. 3.5. We note that 58% of the 1-st rank users successfully receive their desired signals with probability more than 0.6 in the first time slot,
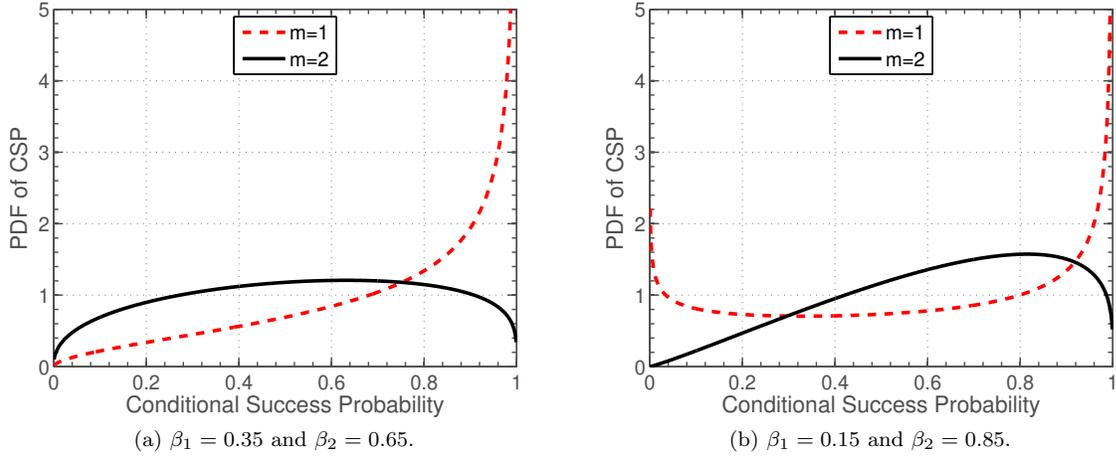
Figure 3.6: PDF of the CSP of 1-st and 2-nd rank downlink NOMA users for $\lambda_{\mathrm{b}} = 0.001$, $N = 2$, $\alpha = 4$, $\theta = -5\,\mathrm{dB}$. (a) $M_{1,(1)} = 0.73$ and $M_{1,(2)} = 0.53$. (b) $M_{1,(1)} = 0.59$ and $M_{1,(2)} = 0.63$

while, after the second time slot, 74% of the 1-st rank users successfully receive their desired signals with probability more than 0.6 (this is obtained by setting $k = 2$ and $x = 0.6$ in (2.6) which yields $\bar{F}_{P_{\mathrm{s},m}}(0.37)$). This value for the 2-nd rank users is 50% and for the 3-rd rank users is 17%. Hence, after the second time slot, 47% of users receive their desired signals with reliability 0.6.

Finite and Infinite Mean Local Delay

Using the beta approximation, distribution of the CSP for the 1-st and 2-nd rank users are shown in Fig. 3.6. To understand the relations between the CSP, the standard success probability (1-st moment), and the mean local delay ($-$1-st moment), consider the following examples.

- When $\lambda_{\mathrm{b}} = 0.001$, $N = 2$, $\alpha = 4$, $\theta = -5\,\mathrm{dB}$, $\beta_1 = 0.35$, and $\beta_2 = 1 - 0.35 = 0.65$, the standard success probability for the 1-st rank users is 0.73 and for the 2-nd rank users is 0.53. For the 1-st and 2-nd rank users, the mean local delays are finite, i.e., $c_m \frac{\delta}{1-\delta} < N - m + 1$ is satisfied for $m = 1$ and $m = 2$.

- When $\beta_1 = 0.15$ and $\beta_2 = 0.85$, the standard success probability for the 1-st rank users is 0.59 and for the 2-nd rank users is 0.63. Although the overall mean success probability for these two examples in Fig. 3.6 is about 0.62, in the second example, for the 1-st rank users, the mean local delay is infinite. When the mean local delay is infinite, it means that there is a significant number of users with small conditional success probabilities in the network [54]. This can also be seen in Fig. 3.6(b) where the PDF of small values of CSP for the 1-st rank users is not zero. Therefore, we can conclude that, in the second example, for the 1-st rank users CSPs are close to 0 and 1 with high probability while for the 2-nd rank users they are close to mean 0.63 with high probability.

37

3.6   Conclusion

Standard success probability is not enough for capturing a network performance since two networks can have the same success probability but completely different mean local delay (for one it can be finite and for the other it can be infinite). On the other hand, meta distribution is a fine-grained performance metric that gives the distribution of the CSP. It can be used in the evaluation of the network performance metrics such as the standard success probability (first positive moment) and mean local delay (first negative moment).

In this chapter, we have developed a stochastic geometry framework to derive the moments of the conditional success probability (CSP) and its meta distribution in uplink and downlink NOMA networks. For uplink NOMA, we have proposed two point process models for spatial locations of the interferers by using the definition of BS/user pair correlation function and demonstrated the accuracy of the models by using Monte-Carlo simulations. For downlink NOMA, we have derived closed-form solutions for the success probability, the meta distribution, and the mean local delay. In the next chapter, we derive the performance of NOMA in a more realistic setup. Specifically, we study the impact of ordering in SIC.

Chapter 4

Accuracy of Distance-Based Ranking of Users in the Analysis of NOMA Systems

We characterize the accuracy of analyzing the performance of a non-orthogonal multiple access (NOMA) system where users are ranked according to their distances instead of instantaneous channel gains, i.e., product of their distance-based path-loss and fading channel gains. Distance-based ranking of users is analytically tractable and can lead to important insights. However, it may not be appropriate in a multipath fading environment where a near user suffers from severe fading while a far user experiences weak fading. Since the ranking of users (and in turn interferers) in a NOMA system has a direct impact on coverage probability analysis, impact of the traditional distance-based ranking, as opposed to instantaneous signal power- (ISP-) based ranking, needs to be understood. This will enable us to identify scenarios where distance-based ranking, which is easier to implement compared to instantaneous signal power-based ranking, is acceptable for system performance analysis. To this end, in this chapter, we derive the probability of the event when distance-based ranking yields the same results as instantaneous signal power-based ranking, which is referred to as the *accuracy probability*. We show that the accuracy probability decreases with the increasing number of users and increases with the path-loss exponent. Closed-form expressions are presented for Rayleigh fading environment.

In addition, through examples, we study the impact of ranking on uplink and downlink coverage probability. In dowlink NOMA, decoding order is dictated by the power allocation at the BS. If BS employs

ISP-based ranking to allocate power, decoding order at each user will be based on users' ISPs (in the same order as the power allocation at the BS), and if BS employs distance-based ranking, decoding order will be based on users' distances (in the same order as the power allocation at the BS). On the other hand, in uplink NOMA, since the channels of different users are different, each message signal experiences distinct channel gain. Therefore, even when user j transmits with more power, compared to user i, it is still possible that the received signal power of user i at the BS be stronger than user j. The strongest signal is decoded first at the BS and experiences interference from all users in the cluster with relatively weaker instantaneous received signal powers (ISPs). Thus, to study the coverage probability for uplink NOMA, we must rank the users based on their received ISPs. In the existing literature, however, mean signal power (MSP)-based ranking[1] is used to determine the decoding order and calculate the SIR coverage probability. Although this approximation provides tractable results, it is not accurate. In this chapter, we also derive the coverage probability for ISP-based ranking, and we show that MSP-based ranking underestimates the SIR coverage probability

4.1   Introduction

Performance of non-orthogonal multiple access (NOMA) in both uplink and downlink depends on the successive intra-cell interference cancellation (SIC) which relies on the ranking of the users in each NOMA cluster [8]. In particular, downlink intra-cell interference received at a given user in NOMA depends on the power allocation factors of users in the cluster. These power allocation factors are designed according to the ranking of users' transmission links quality. For example, users with stronger links have smaller power allocations and vice versa. On the other hand, in uplink NOMA, to apply SIC, BS successively decodes and cancels the messages of strong channel users, prior to decoding the signals of weak channel users [8]. Therefore, the intra-cell interference encountered by any user depends on the instantaneous received signal powers (which includes short-term fading) of users in the NOMA cluster.

The link quality can be evaluated by different metrics. These metrics should include effects of path-loss (and therefore link distance), fading, and/or inter-cell interference [72]. However, acquiring complete channel state information (CSI) with fading and inter-cell interference increases system complexity. Therefore, most of the existing state-of-the-art resorts to mean signal power- (or distance-) based user ranking in NOMA analysis. Recently, in [15], the rate coverage probability of a user at rank m in uplink NOMA has been derived assuming distance-based ranking. In [73, 59], it is assumed that the order statistics of instantaneous signal power are dominated by the distance; hence, in the analysis, users are ordered based on their distances

---

[1]In our network, MSP-based, path-loss-based, and distance-based rankings yield the same result. Thus, we use them interchangeably throughout the chapter.

instead of complete CSI. In [74, 75], distance-based ranking is used for the analysis of NOMA systems with HARQ. In [76], the authors study two-user cooperative NOMA and derive the outage probability assuming the near user to be the strong user and the far user to be the weak user. In [77], a similar assumption is made for the analysis of uplink and downlink MIMO NOMA. In order to maximize the rate region of the uplink NOMA systems, in [78], decoding order of the information signal at the BS is the inverse of the distances. In [72], the authors derive the outage probability in downlink Poisson cellular networks where users are ranked based on mean signal power and instantaneous signal-to-intercell-interference-plus-noise ratio.

To avoid analytical complexity (in theory) and overcome implementation complexity, mean signal power-(distance-) based ranking is typically considered to be appropriate for ordering users in a NOMA cluster. Although this method simplifies the analysis and provides tractable results, its validation (i.e., *accuracy*) has not been studied yet. The distance-based ordering may not always be accurate, especially in a dynamic multipath fading environment, where a near user can experience severe fading and a far user can observe weak fading. Since the ranking of users in a NOMA system has a direct impact on the system performance (e.g., coverage probability) analysis, it is crucial to quantify the impact of distance-based ranking in various environments and to identify the scenarios where this ranking is accurate (i.e., provides system performance close to that achievable with full CSI-based user ranking).

The contributions of this chapter can be summarized as follows:

- This chapter characterizes the accuracy of analyzing the performance of a NOMA system where users are ranked according to their distances (or, equivalently, mean signal powers) instead of instantaneous signal powers, i.e., product of their distance-based path-loss and fading channel. In particular, we derive the probability of the event when distance-based ranking yields same results as instantaneous signal power-based ranking, which is referred to as the *accuracy probability*.

- By analyzing the properties of the derived accuracy probability, we show that the accuracy probability decreases with the increasing number of NOMA users and increases with the path-loss exponent. Closed-form expressions are derived for special cases with two and three users in a NOMA cluster and Rayleigh fading.

- For uplink NOMA, we must rank the users based on their received ISPs. This is unlike downlink NOMA, where the order of decoding at the receiver is dictated by the power allocation at the BS. In this regard, through examples, we further study the impact of ranking on uplink and downlink. We also derive the coverage probability for ISP-based ranking in uplink NOMA.

- Finally, we study the impact of fading severity and user selection on the accuracy probability.

The rest of the chapter is organized as follows. The system model and assumptions are presented in Section 4.2. The definition and properties of the accuracy probability are provided in Section 4.3. Discussions on the impact of ranking on uplink and downlink coverage probability are provided in Section 4.4. For Rayleigh and Nakagami-m fading, the accuracy probability is derived in Sections 4.5 and 4.6, respectively. In Section 4.7, the impact of user pairing on the accuracy probability is investigated. Numerical results are presented in Section 4.8. Finally, Section 4.9 concludes the chapter.

4.2 System Model and Assumptions

We assume that the spatial locations of the BSs follow a homogeneous Poisson point process (PPP) Φ of intensity λ. Users are also distributed according to a homogeneous PPP Φ_{U} of intensity λ_{u} independent of the BS point process. Each user is associated to its nearest BS. We consider a heavily loaded regime, i.e., $\lambda_{\mathrm{u}} \gg \lambda$ where we have at least N users in a typical Voronoi cell[2]. To form a NOMA cluster of size N in the typical cell, we randomly select N users. Therefore, NOMA users are uniformly distributed within the typical Voronoi cell. The explicit distribution of the main geometrical characteristics of the typical cell of a Voronoi tessellation is not known [80]. In [81, 22, 20], taking $c = 5/4$, the probability density function (PDF) and the cumulative distribution function (CDF) of the distance for a typical user from its serving BS are approximated, respectively, as follows:

$$f_r(x) \approx 2c\lambda\pi x e^{-c\lambda\pi x^2}, \qquad F_r(x) \approx 1 - e^{-c\lambda\pi x^2}, \qquad x \geq 0. \tag{4.1}$$

A realization of the network model is illustrated in Fig. 4.1.

Let us denote the distance between the i-th nearest user (termed rank i user) and the serving BS by $r_{(i)}$, $1 \leq i \leq N$. The received power, for the user at rank i, is modeled by $h_i r_{(i)}^{-\alpha}$. $r_{(i)}^{-\alpha}$ represents the large-scale path-loss where $\alpha > 2$ is the path-loss exponent. h_i models the channel power gain due to small-scale fading. The channel power gains follow independent gamma distribution with parameter m and mean Ω for Nakagami-m fading environment, i.e.,

$$f_h(x) = \frac{m^m x^{m-1}}{\Gamma(m)\Omega^m} \exp\left(-\frac{mx}{\Omega}\right), \tag{4.2}$$

where $\Gamma(.)$ is the gamma function. By setting $m = 1$, it reduces to the exponential distribution, corresponding to Rayleigh fading.

---

[2]In a heavily loaded network, when N is small, assuming that we have at least N users in a typical cell, is not unrealistic. For instance, when $\lambda_{\mathrm{u}}/\lambda = 10$, according to [79][Lemma 1], the probability of having more than one user in the typical cell is 0.97 and the probability of having more than two users is 0.93.
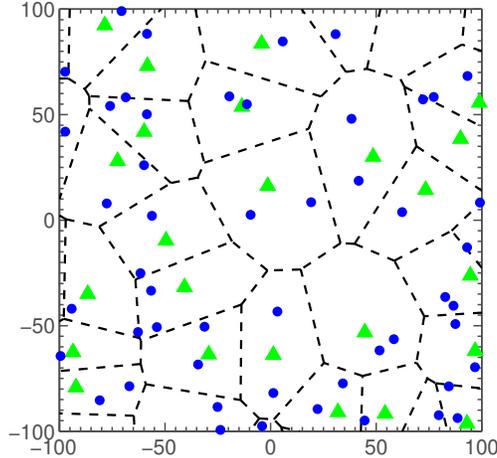
Figure 4.1: A realization of the network. The green triangles represent the BSs and the blue dots represent the users. $\lambda = 0.0005$.

4.3 Ranking Accuracy Probability: Definition and Properties

Ranking users based on their distances from the serving BS in each NOMA cluster is a common assumption in the existing literature to characterize the performance of NOMA. That is, the nearest user to the serving BS is assumed as the user with the highest CSI and so on (which may not always be true). To understand the accuracy of this approximation and its impact on important performance metrics such as coverage probability, in this and next sections, we define the term *(ranking) accuracy probability* $\mathcal{A}$, highlight its properties, and describe its connection to uplink and downlink coverage probability.

Definition 4.1 (Ranking Accuracy probability). *Ranking Accuracy probability $\mathcal{A}$ is the probability that ordering based on large-scale path loss[3] matches ordering based on the instantaneous signal power (small-scale fading and large-scale path-loss), i.e.,*

$$
\begin{aligned}
\mathcal{A} &= \mathbb{P}\left(h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} > \cdots > h_N r_{(N)}^{-\alpha} \mid r_{(1)} < r_{(2)} < \cdots < r_{(N)}\right) \\
&= \mathbb{P}\left(h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} > \cdots > h_N r_{(N)}^{-\alpha}\right),
\end{aligned}
\tag{4.3}
$$

where the last result is obtained from the definition of $r_{(i)}$. According to the definition, ordering users based on path-loss, instead of instantaneous signal power is accurate with probability $\mathcal{A}$.

    Example: According to the definition of the accuracy probability, in two-UE uplink NOMA, received signal power at the BS from the near user is stronger than that for the far user with probability $\mathcal{A}$. With probability $1 - \mathcal{A}$ received signal power for the far user is stronger.

---

[3]In our network, MSP-based, path-loss-based, and distance-based rankings are the same so we use them interchangeably throughout the chapter.

Using the indicator function, the ranking accuracy probability can be expressed as

$$
\begin{aligned}
\mathcal{A} &= \mathbb{E}_{\{h_i\},\{r_{(i)}\}} \left[\mathbf{1} \left(h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} > \cdots > h_N r_{(N)}^{-\alpha} \right) \right] \\
&= \mathbb{E}_{\{r_{(i)}\}} \left[\mathbb{E}_{\{h_i\}} \left[\mathbf{1} \left(h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} > \cdots > h_N r_{(N)}^{-\alpha} \right) \right] \right].
\end{aligned}
\tag{4.4}
$$

The inner expectation in (4.4) is over the channel power gains $\{h_i\}$, i.e., the inner expectation calculates the ranking accuracy probability for a given realization of users and BSs. The outer expectation is with respect to the ordered desired link distances $\{r_{(i)}\}$. In the derivation of the outer expectation we use the following definition.

Definition 4.2 (Joint PDF of N-ordered Random Variables)**.** *Let $r_1, r_2, ..., r_N$ be a set of N i.i.d. random variables with PDF $f_r(x)$. Let $r_{(i)}$ denote the i-th smallest observation of the N random variables, i.e., $r_{(1)} \leq r_{(2)} \leq \cdots \leq r_{(N)}$. The joint PDF of N-ordered random variables is given as [64]:*

$$
f_{r_{(1)}, \cdots, r_{(N)}}(x_1, \cdots, x_N) = \begin{cases} N! \prod_{i=1}^{N} f_r(x_i) & \text{when } x_1 \leq x_2 \leq \cdots \leq x_N \\ 0 & \text{otherwise} \end{cases}.
\tag{4.5}
$$

In the following, two properties of the ranking accuracy probability are reported. These properties are general and apply to any fading channel and users' spatial distributions.

Corollary 4.1. *The ranking accuracy probability $\mathcal{A}$ fulfills the following properties: i) $\mathcal{A}$ is a decreasing function of NOMA cluster size N, ii) $\mathcal{A}$ is an increasing function of path-loss exponent α.*

Proof. The result in (i) follows from definition of the ranking accuracy probability (4.3). The result in (ii) follows from (4.4). If $h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} > \cdots > h_N r_{(N)}^{-\alpha}$ is satisfied by α, it will also be satisfied by higher values of path-loss exponent. On the other hand, if $h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} > \cdots > h_N r_{(N)}^{-\alpha}$ is not satisfied by α, any smaller value of path-loss exponent cannot also satisfy this condition. Therefore, when we increase the path-loss exponent, ranking users based on their distances is valid for a wider range of channel and distance realizations, i.e., $\mathcal{A}$ is an increasing function of α. $\square$

4.4 Ranking in Uplink and Downlink NOMA

Uplink NOMA: All users of the same NOMA cluster transmit in the same time, frequency, and code domain. Since the signal of each user experiences a different fading and path-loss, at the BS we have

different signal powers. To apply SIC, in the first step, the BS decodes the strongest signal by treating other signals as noise. Then, it remodulates the decoded signal and subtracts the remodulated signal from the composite received signal and proceeds the SIC process on the remaining signal by decoding, regenerating, and canceling the second strongest signal and so on [15]. Now consider the typical cell and let us denote transmit signals of near and far users by x_1 and x_2, respectively, and assume they have unit powers. The received signal at the BS from these two users is $\sqrt{h_1 r_{(1)}^{-\alpha}} x_1 + \sqrt{h_2 r_{(2)}^{-\alpha}} x_2$. Since uplink inter-cell interference at the desired BS is same for all users in the NOMA cluster, ranking users based on instantaneous signal-to-interference-plus-noise ratio yields the same result as ranking users based on instantaneous signal power. Therefore, x_1 is decoded first when $h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha}$; otherwise, x_2 is decoded first. Thus, for the 2-UE NOMA, the coverage probability of the near user to the BS ($P_{\text{cov},(1)}^{\text{ISP}}$) must be derived as follows:

$$
\begin{aligned}
P_{\text{cov},(1)}^{\text{ISP}} = \mathbb{P} & \left\{ \frac{P_{\text{tx}} h_1 r_{(1)}^{-\alpha}}{P_{\text{tx}} h_2 r_{(2)}^{-\alpha} + I_{\text{inter}} + \sigma_n^2} > \theta \mid h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} \right\} \underbrace{\mathbb{P} \left(h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} \right)}_{\mathcal{A}} \\
& + \mathbb{P} \left\{ \frac{P_{\text{tx}} h_2 r_{(2)}^{-\alpha}}{P_{\text{tx}} h_1 r_{(1)}^{-\alpha} + I_{\text{inter}} + \sigma_n^2} > \theta, \frac{P_{\text{tx}} h_1 r_{(1)}^{-\alpha}}{I_{\text{inter}} + \sigma_n^2} > \theta \mid h_1 r_{(1)}^{-\alpha} < h_2 r_{(2)}^{-\alpha} \right\} \mathbb{P} \left(h_1 r_{(1)}^{-\alpha} < h_2 r_{(2)}^{-\alpha} \right), \quad (4.6)
\end{aligned}
$$

where I_{inter} denotes the inter-cell interference; σ_n^2 is the noise power, and P_{tx} is the transmit power (for proof refer to **Appendix B.1**). According to (4.6), when $h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha}$, BS decodes the intended signal of near user in the presence of interference from far user, and when $h_1 r_{(1)}^{-\alpha} < h_2 r_{(2)}^{-\alpha}$, BS decodes and cancels the signal of far user and then decodes the intended signal of near user[4]. Similarly, for the far user, we have

$$
\begin{aligned}
P_{\text{cov},(2)}^{\text{ISP}} = \mathbb{P} & \left\{ \frac{P_{\text{tx}} h_1 r_{(1)}^{-\alpha}}{P_{\text{tx}} h_2 r_{(2)}^{-\alpha} + I_{\text{inter}} + \sigma_n^2} > \theta, \frac{P_{\text{tx}} h_2 r_{(2)}^{-\alpha}}{I_{\text{inter}} + \sigma_n^2} > \theta \mid h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} \right\} \mathbb{P} \left(h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} \right) \\
& + \mathbb{P} \left\{ \frac{P_{\text{tx}} h_2 r_{(2)}^{-\alpha}}{P_{\text{tx}} h_1 r_{(1)}^{-\alpha} + I_{\text{inter}} + \sigma_n^2} > \theta \mid h_1 r_{(1)}^{-\alpha} < h_2 r_{(2)}^{-\alpha} \right\} \mathbb{P} \left(h_1 r_{(1)}^{-\alpha} < h_2 r_{(2)}^{-\alpha} \right). \quad (4.7)
\end{aligned}
$$

In the analysis of uplink NOMA, it is generally assumed that the nearest user to the BS has the highest instantaneous signal power, i.e., $\mathbb{P} \left(h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} \right) \approx 1$ for 2-UE NOMA. Hence, only the first terms in (4.6) and (4.7) are derived to date in the literature and reported as $P_{\text{cov},(1)}^{\text{ISP}}$ and $P_{\text{cov},(2)}^{\text{ISP}}$, respectively. However, the first terms in (4.6) and (4.7) provide good approximations when: i) network is intercell-interference- (or noise-) limited, or ii) assumption $\mathbb{P} \left(h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} \right) \approx 1$ is accurate.

Downlink NOMA: The BS allocates different powers to different users of the NOMA cluster, and then transmits the superposed signal. At the receiver side, similar to uplink, each user employs the SIC process to decode its intended signal. As an example, consider 2-UE downlink NOMA. The BS transmits the

---

[4]This corresponds to the worst-case SIC, where to decode a weak user's signal we must successfully decode stronger users' signals first.

superposed signal $\sqrt{a_1 P_{\mathrm{BS}}}x_1 + \sqrt{a_2 P_{\mathrm{BS}}}x_2$, where $a_1 P_{\mathrm{BS}}$ and $a_2 P_{\mathrm{BS}}$ are the allocated powers to the near and far users, respectively. The received signal from the serving BS at user i, $i = 1$ for near user and $i = 2$ for far user, is $\sqrt{h_i r_{(i)}^{-\alpha}}(\sqrt{a_1 P_{\mathrm{BS}}}x_1 + \sqrt{a_2 P_{\mathrm{BS}}}x_2)$. Order of decoding at a given user depends on the power allocations of users' signals at the BS since effect of fading and path-loss is the same on all signals[5]. If BS allocates more power to the far user (i.e., $a_2 > a_1$), far user decodes its intended signal in the presence of interference from the near user while near user decodes far user's signal before decoding its intended signal. Typically BS allocates more power to the weak user. To determine the weak user, knowledge of complete CSI with fading and interference is required. However, acquiring complete CSI increases system complexity. Therefore, in the following, we consider instantaneous signal power-based and distance-based ranking. Let denote the allocated powers to the weak and strong users by $a_{(2)} P_{\mathrm{BS}}$ and $a_{(1)} P_{\mathrm{BS}}$, where $0 < a_{(1)} < a_{(2)} < 1$ and $a_{(1)} + a_{(2)} = 1$. With instantaneous signal power-based ranking at the BS, the coverage probabilities of near and far users are as follows:

$$
\begin{aligned}
P_{\mathrm{cov},(1)}^{\mathrm{ISP}} = {} & \mathbb{P}\left\{ \frac{a_{(2)} P_{\mathrm{BS}} h_1 r_{(1)}^{-\alpha}}{a_{(1)} P_{\mathrm{BS}} h_1 r_{(1)}^{-\alpha} + I_{\mathrm{inter}}^{(1)} + \sigma_n^2} > \theta, \; \frac{a_{(1)} P_{\mathrm{BS}} h_1 r_{(1)}^{-\alpha}}{I_{\mathrm{inter}}^{(1)} + \sigma_n^2} > \theta \mid h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} \right\} \mathbb{P}\left(h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} \right) \\
& + \mathbb{P}\left\{ \frac{a_{(2)} P_{\mathrm{BS}} h_1 r_{(1)}^{-\alpha}}{a_{(1)} P_{\mathrm{BS}} h_1 r_{(1)}^{-\alpha} + I_{\mathrm{inter}}^{(1)} + \sigma_n^2} > \theta \mid h_1 r_{(1)}^{-\alpha} < h_2 r_{(2)}^{-\alpha} \right\} \mathbb{P}\left(h_1 r_{(1)}^{-\alpha} < h_2 r_{(2)}^{-\alpha} \right),
\end{aligned}
\tag{4.8}
$$

$$
\begin{aligned}
P_{\mathrm{cov},(2)}^{\mathrm{ISP}} = {} & \mathbb{P}\left\{ \frac{a_{(2)} P_{\mathrm{BS}} h_2 r_{(2)}^{-\alpha}}{a_{(1)} P_{\mathrm{BS}} h_2 r_{(2)}^{-\alpha} + I_{\mathrm{inter}}^{(2)} + \sigma_n^2} > \theta \mid h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} \right\} \mathbb{P}\left(h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} \right) \\
& + \mathbb{P}\left\{ \frac{a_{(2)} P_{\mathrm{BS}} h_2 r_{(2)}^{-\alpha}}{a_{(1)} P_{\mathrm{BS}} h_2 r_{(2)}^{-\alpha} + I_{\mathrm{inter}}^{(2)} + \sigma_n^2} > \theta, \; \frac{a_{(1)} P_{\mathrm{BS}} h_2 r_{(2)}^{-\alpha}}{I_{\mathrm{inter}}^{(2)} + \sigma_n^2} > \theta \mid h_1 r_{(1)}^{-\alpha} < h_2 r_{(2)}^{-\alpha} \right\} \mathbb{P}\left(h_1 r_{(1)}^{-\alpha} < h_2 r_{(2)}^{-\alpha} \right).
\end{aligned}
\tag{4.9}
$$

$I_{\mathrm{inter}}^{(1)}$ and $I_{\mathrm{inter}}^{(2)}$ denote the inter-cell interference at the near and far users, respectively. With instantaneous signal power-based ranking at the BS, when $h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha}$, BS allocates more power to the far user, while, when $h_1 r_{(1)}^{-\alpha} < h_2 r_{(2)}^{-\alpha}$, more power is allocated to the near user. On the other hand, with distance-based ranking at the BS, BS always allocates more power to the far user, i.e., far user is always considered as the weak user. Therefore, coverage probabilities of the near and far users with distance-based ranking at the BS can be derived by the first terms in (4.8) and (4.9), respectively. Note that coverage probabilities with distance-based ranking provide close results to the coverage probabilities with instantaneous signal power-based ranking when: i) the network is intracell-interference-limited, or ii) the assumption

---

[5]In the downlink, for a fixed power allocation at the transmitter (BS), decoding order is fixed at the receiver (user) side since, at each receiver, effect of channel is the same on all signals. Therefore, in downlink, decoding order is completely determined by the power allocation at the BS. On the other hand, in uplink, different signals experience different channels thus decoding order at the receiver (BS) may change with channel conditions, even for a fixed power allocation at the transmitter (user) side.

$\mathbb{P}\left(h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha}\right) \approx 1$ is accurate.

Comparison: In dowlink NOMA, decoding order is dictated by the power allocation at the BS. If BS employs ISP-based ranking to allocate power, decoding order at each user will be based on users' ISPs (in the same order as the power allocation at the BS), and if the BS employs MSP-based ranking, decoding order will be based on users' MSPs (in the same order as the power allocation at the BS). Thus, derived analytical results for downlink with MSP-based ranking, in the literature, are valid when the BS employs MSP-based ranking for power allocation.

On the other hand, in uplink NOMA, since the channels of different users are different, each message signal experiences distinct channel gain. Therefore, even when user j transmits with more power, compared to user i, it is still possible that the received signal power of user i at the BS be stronger than user j. The strongest signal is decoded first at the BS and experiences interference from all users in the cluster with relatively weaker ISPs. Thus, to study the coverage probability for uplink NOMA, we must rank the users based on their received ISPs. In the existing literature, however, MSP-based ranking is used to determine the decoding order and calculate the uplink coverage probability. Since this is not accurate, we have derived the uplink coverage probability with ISP-based ranking in **Appendix B.2**.

4.5   Accuracy Probability for Rayleigh Fading

In this section, we derive the (ranking) accuracy probability $\mathcal{A}$ for Rayleigh fading. Note that the results provided in this section for Rayleigh fading can also be obtained from the results derived in the next section for Nakagami-m fading by setting $m = 1$. However, for N-UE NOMA, calculating inner expectation in (4.4) for Nakagami-m fading yields $N - 1$ integrals as is shown in **Theorem 4.3**. Deriving $\mathcal{A}$ for Rayleigh fading directly from **Definition 4.1** is easier than deriving from **Theorem 4.3** except for some special cases such as $N = 2$. Therefore, we first study the accuracy probability for Rayleigh fading in this section. Since steps of the proofs for Rayleigh and Nakagami-m fading are similar, in the next section, we will only mention the steps or directly provide the final expressions. It is worth mentioning that the result in **Theorem 4.2** can be obtained by solving the integrals in **Theorem 4.4** for $m = 1$ and applying binomial expansion. Our methodology to derive $\mathcal{A}$ can be described as follows:

1. Derive the inner expectation in (4.4) by averaging over fading channel powers $\{h_i\}$.

2. Characterize $\mathcal{A}$ for any arbitrary users' location model.

3. Derive $\mathcal{A}$ by averaging over the distance distribution of users considering PPP model.

The first two steps are performed in the following theorem and the third step is conducted in **Theorem 4.2** and in subsequent discussions.

Theorem 4.1 (Accuracy Probability of Distance-Based Approximation - N UE NOMA)**.** *For N-UE NOMA, the inner expectation in* (4.4) *for Rayleigh fading can be obtained by*

$$\mathbb{E}_{\{h_i\}}\left[\mathbf{1}\left(h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} > \cdots > h_N r_{(N)}^{-\alpha}\right)\right] = \prod_{i=2}^{N} \frac{1}{\sum_{j=1}^{i}\left(\frac{r_{(j)}}{r_{(i)}}\right)^{\alpha}}. \tag{4.10}$$

Then, using aforementioned expression, (4.10), *and* ***Definition 4.2*** *we obtain*

$$\mathcal{A} = \mathbb{E}\left[\prod_{i=2}^{N}\frac{1}{\sum_{j=1}^{i}\left(\frac{r_{(j)}}{r_{(i)}}\right)^{\alpha}}\right]$$

$$= N! \int_0^\infty \int_{r_1}^\infty \cdots \int_{r_{N-1}}^\infty \prod_{i=2}^{N}\frac{1}{\sum_{j=1}^{i}\left(\frac{r_j}{r_i}\right)^{\alpha}} f_r(r_1) f_r(r_2)\cdots f_r(r_N)\mathrm{d}r_N\cdots\mathrm{d}r_2\mathrm{d}r_1. \tag{4.11}$$

Proof. See **Appendix B.3**. □

According to **Theorem 4.1**, the accuracy probability does not depend on the mean channel power gain Ω. In the following corollary, we provide a simplified expression for the inner expectation in (4.4) considering 2-UE NOMA cluster[6].

Corollary 4.2. *Substituting $N = 2$ in* ***Theorem 4.1*** *and using negative binomial series* $(1 + x)^{-1} = \sum_{k=0}^\infty(-1)^k x^k$ *for* $|x| < 1.$*, the inner expectation in* (4.4) *for Rayleigh fading can be obtained as follows:*

$$\mathbb{E}_{\{h_i\}}\left[\mathbf{1}\left(h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha}\right)\right] = \sum_{k=0}^\infty (-1)^k \left(\frac{r_{(1)}}{r_{(2)}}\right)^{\alpha k}. \tag{4.12}$$

The accuracy probability can then be derived as follows:

$$\mathcal{A} = \sum_{k=0}^\infty (-1)^k \mathbb{E}\left[\left(\frac{r_{(1)}}{r_{(2)}}\right)^{\alpha k}\right] = 2\sum_{k=0}^\infty (-1)^k \int_0^\infty \int_{r_1}^\infty \left(\frac{r_1}{r_2}\right)^{\alpha k} f_r(r_1) f_r(r_2)\mathrm{d}r_2\mathrm{d}r_1. \tag{4.13}$$

It is worth mentioning that the summation in (4.13) can be truncated after a few terms since the expression inside the summation is close to zero for large values of k. Moreover, unlike the expectation in (4.11), the expectation in (4.13) can be derived in closed-form for PPP. In the following, we obtain the accuracy probability in closed-form considering two users in a NOMA cluster (i.e., 2-UE NOMA), and then

---

[6]In this chapter, we use the term "N-UE NOMA" to make it explicit that the framework can capture any value of N; however, the performance gains of NOMA over OMA (Orthogonal Multiple Access) are generally achievable for small number of user equipment (UE) in a NOMA cluster. Therefore, we are more interested in cases where $N = 2$ and $N = 3$.

we study the accuracy probability for N-UE NOMA. Evidently, for N-UE NOMA, there is no closed-form expression available.

Theorem 4.2 (Accuracy Probability of Distance-Based Approximation - 2 UE NOMA and PPP Model)**.** *When each BS serves users that are located in its Voronoi cell, the accuracy probability for 2-UE NOMA with Rayleigh fading is*

$$\mathcal{A} = \sum_{k=0}^{\infty} \frac{(-1)^k}{\alpha k + 2} \, {}_2F_1\left(2, 1; \frac{\alpha k}{2} + 2; \frac{1}{2}\right).$$

Proof. See **Appendix B.4**. $\qquad\square$

According to **Theorem 4.2**, for $N = 2$, the probability that the path-loss-based ranking matches the instantaneous signal power-based ranking only depends on path-loss exponent α and does not depend on BS intensity λ. Similarly, we can generalize **Theorem 4.2** to the case of N-UE NOMA as stated in the following corollary.

Corollary 4.3. *For N-UE NOMA with Rayleigh fading, the accuracy probability $\mathcal{A}$ only depends on path-loss exponent α and does not depend on BS intensity λ.*

Proof. See **Appendix B.5**. $\qquad\square$

4.6   Accuracy Probability for Nakagami-m Fading

In this section, we derive the probability of accuracy of distance-based approximation considering Nakagami-m fading channels. Using the analytical results, in Section 4.8, we will show that for more severe fading conditions (i.e., for small values of m), the distance-based approximation is less accurate whereas for higher values of m, the distance-based approximation is more accurate.

Similar to the previous subsection, we first derive the inner expectation in (4.4). Then the accuracy probability $\mathcal{A}$ is obtained for PPP model.

Theorem 4.3 (Probability of the Accuracy of Distance-based Approximation - N-UE NOMA)**.** *For Nakagami-m fading, with shape parameter m, the inner expectation in (4.4) can be derived as follows:*

$$\mathbb{E}_{\{h_i\}}\left[\mathbf{1}\left(h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} > \cdots > h_N r_{(N)}^{-\alpha}\right)\right]$$
$$= \frac{\Gamma(Nm)}{\Gamma(m)^N} \int_1^{\infty} \int_1^{\infty} \cdots \int_1^{\infty} \frac{1}{\left[1 + \sum_{i=1}^{N-1}\left(\frac{r_{(i)}}{r_{(N)}}\right)^{\alpha} \prod_{k=i}^{N-1} t_k\right]^{Nm}} \prod_{j=1}^{N-1}\left(\frac{r_{(j)}}{r_{(N)}}\right)^{\alpha m} t_j^{jm-1} \mathrm{d}t_{N-1} \cdots \mathrm{d}t_2 \mathrm{d}t_1.$$

*Then $\mathcal{A}$ can be derived by averaging over the desired link distance distribution using **Definition 4.2**.*

Proof. See **Appendix B.6**. □

Similar to Rayleigh fading, the accuracy probability for Nakagami-m fading does not depend on mean channel power gain. By setting $N = 2$ for 2-UE NOMA, we obtain

$$\mathbb{E}_{\{h_i\}}\left[\mathbf{1}\left(h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha}\right)\right] = \frac{\Gamma(2m)}{\Gamma(m)^2}\left(\frac{r_{(1)}}{r_{(2)}}\right)^{\alpha m}\int_1^\infty \frac{t_1^{m-1}\mathrm{d}t_1}{\left[1 + \left(\frac{r_{(1)}}{r_{(2)}}\right)^\alpha t_1\right]^{2m}}. \tag{4.14}$$

For 2-UE NOMA, when $m = 1$ (Rayleigh fading), (4.14) reverts to **Corollary 4.2**. However, deriving **Theorem 4.1** from **Theorem 4.3** for N-UE NOMA, when $m = 1$, is not straightforward. Using **Definition 4.2** with (4.14) for 2-UE NOMA and with **Theorem 4.3** for N-UE NOMA provides the accuracy probability.

Theorem 4.4 (Probability of the Accuracy of Distance-based Approximation - 2 UE NOMA and PPP Model)**.** *For 2-UE NOMA and PPP model for users' spatial locations, the accuracy probability for Nakagami-m fading with fading parameter m is as follows:*

$$\mathcal{A} = \frac{4\Gamma(2m)}{\Gamma(m)\Gamma(m+1)}\int_0^1 \frac{u^{1-\alpha m}}{(1+u^2)^2}\,_2F_1(2m, m; m+1; -u^{-\alpha})\mathrm{d}u.$$

Proof. The accuracy probability can be derived as:

$$\mathcal{A} = \frac{2\Gamma(2m)}{\Gamma(m)^2}(2c\lambda\pi)^2\int_1^\infty\int_0^\infty\int_{r_1}^\infty \frac{\left(\frac{r_1}{r_2}\right)^{\alpha m}t_1^{m-1}}{\left[1 + \left(\frac{r_1}{r_2}\right)^\alpha t_1\right]^{2m}}r_1 r_2 e^{-c\lambda\pi(r_1^2+r_2^2)}\mathrm{d}r_2\mathrm{d}r_1\mathrm{d}t_1$$

$$\overset{(a)}{=} \frac{2\Gamma(2m)}{\Gamma(m)^2}(2c\lambda\pi)^2\int_1^\infty\int_0^1 \frac{u^{\alpha m-3}t_1^{m-1}}{[1+u^\alpha t_1]^{2m}}\int_0^\infty v^3 e^{-c\lambda\pi\left(1+\frac{1}{u^2}\right)v^2}\mathrm{d}v\mathrm{d}u\mathrm{d}t_1$$

$$\overset{(b)}{=} \frac{4\Gamma(2m)}{\Gamma(m)^2}\int_1^\infty\int_0^1 \frac{u^{\alpha m+1}t_1^{m-1}}{[1+u^\alpha t_1]^{2m}}\frac{1}{(1+u^2)^2}\mathrm{d}u\mathrm{d}t_1$$

$$\overset{(c)}{=} \frac{4\Gamma(2m)}{\Gamma(m)^2}\int_0^1 \frac{u^{1-\alpha m}}{(1+u^2)^2}\int_0^1 \frac{z^{m-1}}{[1+u^{-\alpha}z]^{2m}}\mathrm{d}z\mathrm{d}u,$$

where (a) is obtained by changes of variables $\frac{r_1}{r_2} = u$ and $r_1 = v$. (b) follows by applying $c\lambda\pi\left(1+\frac{1}{u^2}\right)v^2 = x$. (c) is obtained by $t_1^{-1} = z$. Finally, **Theorem 4.4** is derived by using the integral representation of Gaussian hypergeometric function. □

According to **Theorem 4.4**, the accuracy probability in Nakagami-m fading for 2-UE NOMA does not depend on the BS intensity λ. In the following, we prove that, for N-UE NOMA with Nakagami-m fading, the accuracy probability is independent of λ.

Corollary 4.4. *For N-UE NOMA and PPP model, the accuracy probability for Nakagami-m fading with parameter m is independent of the BS intensity λ.*

Proof. See **Appendix B.7**. □

4.7 User Pairing and Accuracy Probability

In the previous sections, from the set of users that are associated to the same BS, N users were randomly selected to form a NOMA cluster. However, in practice, NOMA users are chosen such that NOMA gain can be achieved over OMA. For instance, to form a 2-UE NOMA cluster, out of M users associated to the typical BS, usually the nearest and the farthest users are selected. In the following, we study the accuracy probability with user pairing.

To form a NOMA cluster, we have selected N users from M users that are associated to the typical BS. We denote rank of the selected users by the set $s = \{s_{(i)}\}$, where $i = 1, 2, \cdots, N$, $s_{(i)} \in \{1, 2, ..., M\}$, and $1 \leq s_{(1)} < s_{(2)} \cdots < s_{(N-1)} < s_{(N)} \leq M$. From **Definition 4.2** and **Theorem 4.1**, for Rayleigh fading, we obtain

$$
\mathcal{A} = \mathbb{E}\left[\prod_{i=2}^{N} \frac{1}{\sum_{j=1}^{i} \left(\frac{r_{s_{(j)}}}{r_{s_{(i)}}}\right)^{\alpha}}\right]
$$

$$
= M! \int_0^\infty \int_{r_1}^\infty \cdots \int_{r_{M-1}}^\infty \prod_{i=2}^{N} \frac{1}{\sum_{j=1}^{i} \left(\frac{r_{s_{(j)}}}{r_{s_{(i)}}}\right)^{\alpha}} f_r(r_1) f_r(r_2) \cdots f_r(r_M) \mathrm{d}r_M \cdots \mathrm{d}r_2 \mathrm{d}r_1. \tag{4.15}
$$

For $N = 2$, when we select the nearest and the farthest user, i.e, $s_{(1)} = 1$ and $s_{(2)} = M$, (4.15) can be simplified as in the following:

$$
\mathcal{A} = M! \int_0^\infty \int_{r_1}^\infty \cdots \int_{r_{M-1}}^\infty \frac{1}{1 + \left(\frac{r_1}{r_M}\right)^{\alpha}} f_r(r_1) f_r(r_2) \cdots f_r(r_M) \mathrm{d}r_M \cdots \mathrm{d}r_2 \mathrm{d}r_1
$$

$$
\overset{(a)}{=} \frac{M!}{(M-2)!} \int_0^\infty \int_{r_1}^\infty \frac{1}{1 + \left(\frac{r_1}{r_M}\right)^{\alpha}} \left[F_r(r_M) - F_r(r_1)\right]^{M-2} f_r(r_1) f_r(r_M) \mathrm{d}r_M \mathrm{d}r_1, \tag{4.16}
$$

where (a) is obtained using the technique in [82] to derive Equation 2.12, i.e., for i.i.d. random variables $r_2, r_3, \cdots, r_{M-1}$, $\left[F_r(r_M) - F_r(r_1)\right]^{M-2}$ is the probability that they are in the interval $[r_1, r_M]$. Sorting these random variables in an ascending order based on their realizations gives $(M-2)!$ different permutations out of which only one satisfies the condition $r_2 < r_3 < \cdots < r_{M-1}$. We can similarly simplify (4.15) for other values of N and different selection of NOMA users. Note that the same result can also be obtained by

averaging the result in **Theorem 4.1** with respect to the joint PDF of $r_{s_{(1)}}, r_{s_{(2)}}, \cdots, r_{s_{(N)}}$, which is also provided in [82].

Corollary 4.5. *For 2-UE NOMA, the accuracy probability, when we select the nearest and the farthest user, i.e., $s_{(1)} = 1$ and $s_{(2)} = M$, is an increasing function of M irrespective of the fading channel and users' spatial distributions.*

Proof. See **Appendix B.8**. $\qquad\square$

For Nakagami-m fading, using **Definition 4.2** and **Theorem 4.3** gives

$$\mathcal{A} = \frac{\Gamma(Nm)}{\Gamma(m)^N} \int_1^\infty \int_1^\infty \cdots \int_1^\infty \mathbb{E}\left[\frac{\prod_{j=1}^{N-1}\left(\frac{r_{s_{(j)}}}{r_{s_{(N)}}}\right)^{\alpha m}}{\left[1 + \sum_{i=1}^{N-1}\left(\frac{r_{s_{(i)}}}{r_{s_{(N)}}}\right)^\alpha \prod_{k=i}^{N-1} t_k\right]^{Nm}} \right] \prod_{j=1}^{N-1} t_j^{jm-1} \mathrm{d}t_{N-1}\cdots\mathrm{d}t_2\mathrm{d}t_1, \quad (4.17)$$

where

$$\mathbb{E}\left[\frac{\prod_{j=1}^{N-1}\left(\frac{r_{s_{(j)}}}{r_{s_{(N)}}}\right)^{\alpha m}}{\left[1 + \sum_{i=1}^{N-1}\left(\frac{r_{s_{(i)}}}{r_{s_{(N)}}}\right)^\alpha \prod_{k=i}^{N-1} t_k\right]^{Nm}} \right] =$$

$$M! \int_0^\infty \int_{r_1}^\infty \cdots \int_{r_{M-1}}^\infty \frac{\prod_{j=1}^{N-1}\left(\frac{r_{s_{(j)}}}{r_{s_{(N)}}}\right)^{\alpha m}}{\left[1 + \sum_{i=1}^{N-1}\left(\frac{r_{s_{(i)}}}{r_{s_{(N)}}}\right)^\alpha \prod_{k=i}^{N-1} t_k\right]^{Nm}} f_r(r_1) f_r(r_2)\cdots f_r(r_M)\mathrm{d}r_M\cdots\mathrm{d}r_2\mathrm{d}r_1. \quad (4.18)$$

Now using the above equations, we can study $\mathcal{A}$ for PPP model.

Corollary 4.6. *When each BS serves users in its Voronoi cell (PPP model), for any selection of users for the NOMA cluster, the accuracy probability is independent of the BS intensity λ.*

Proof. The proof can be obtained by using the same approach as in the proof of **Corollary 4.3** for (4.15) (Rayleigh fading) and (4.18) (for Nakagami-m fading). $\qquad\square$

4.8 Numerical and Simulation Results

This section demonstrates the efficacy of the derived expressions by comparing them to Monte-Carlo simulations. In Table 4.1, we provide the final expressions for the accuracy of the distance-based approximation in NOMA assuming PPP, MCP, and TCP spatial models. In the MCP model, around each BS, N NOMA users are uniformly distributed within distance R. In TCP, users are independently and identically distributed

following a normal distribution with variance σ^2 around each BS. Although we have only provided the analytical results for the PPP model in this chapter, our framework can be easily extended to MCP and TCP models.

We use Gaussian quadrature method to approximate and solve four or higher dimensional integrals in Table 4.1. In the following, we briefly review the Gaussian quadrature method, describe simulation parameters, and then present our results which demonstrate the impact of path-loss exponent, fading parameter m (in Nakagami-m fading), and user pairing on the accuracy probability.

Table 4.1: Accuracy Probability $\mathcal{A}$ for Random User Selection

Fading	**Netwrok Model**	N	**Accuracy Probability ($\mathcal{A}$)**
Rayleigh	PPP/TCP	2	$\sum_{k=0}^{\infty} \frac{(-1)^k}{2+\alpha k}\, {}_2F_1\left(2,1;\frac{\alpha k}{2}+2;\frac{1}{2}\right)$
Rayleigh	PPP/TCP	3	$48 \int_0^1 \int_0^1 \frac{u_1 u_2^3}{\left(1+u_1^\alpha\right)\left(1+u_2^\alpha+u_1^\alpha u_2^\alpha\right)\left(1+u_2^2+u_1^2 u_2^2\right)^3} \mathrm{d}u_2 \mathrm{d}u_1$
Raylegih	MCP	2	$\sum_{k=0}^{\infty}(-1)^k \frac{2}{2+\alpha k}$
Rayleigh	MCP	3	$8 \int_0^1 \int_0^1 \frac{u_1 u_2^3}{\left(1+u_1^\alpha\right)\left(1+u_2^\alpha+u_1^\alpha u_2^\alpha\right)} \mathrm{d}u_2 \mathrm{d}u_1$
Nakagami	PPP/TCP	2	$\frac{4\Gamma(2m)}{\Gamma(m)\Gamma(m+1)} \int_0^1 \frac{u^{1-\alpha m}}{(1+u^2)^2}\, {}_2F_1(2m,m;m+1;-u^{-\alpha})\mathrm{d}u$
Nakagami	PPP/TCP	3	$\frac{48\Gamma(3m)}{\Gamma(m)^3} \int_0^1 \int_0^1 \int_0^1 \int_0^1 \frac{u_1^{1+\alpha m} u_2^{3+2\alpha m} z_1^{2m-1} z_2^{m-1}}{\left(u_1^\alpha u_2^\alpha+u_2^\alpha z_1+z_1 z_2\right)^{3m}\left(1+u_2^2+u_1^2 u_2^2\right)^3} \mathrm{d}u_2 \mathrm{d}u_1 \mathrm{d}z_2 \mathrm{d}z_1$
Nakagami	MCP	2	$\frac{2\Gamma(2m)}{\Gamma(m)\Gamma(m+1)} \int_0^1 u^{1-\alpha m}\, {}_2F_1(2m,m;m+1;-u^{-\alpha})\mathrm{d}u$
Nakagami	MCP	3	$\frac{8\Gamma(3m)}{\Gamma(m)^3} \int_0^1 \int_0^1 \int_0^1 \int_0^1 \frac{u_1^{1+\alpha m} u_2^{3+2\alpha m} z_1^{2m-1} z_2^{m-1}}{\left(u_1^\alpha u_2^\alpha+u_2^\alpha z_1+z_1 z_2\right)^{3m}} \mathrm{d}u_2 \mathrm{d}u_1 \mathrm{d}z_2 \mathrm{d}z_1$

4.8.1 Approximation of Multi-Dimensional Integrals

A quadrature rule provides an approximation of the definite integral of a function, usually stated as a weighted sum of function values at specified points within the domain of integration.

Definition 4.3 (Gaussian Quadrature). *When domain of integration is $[0,1]$[7], an n-point Gaussian quadrature rule states*

$$\int_0^1 f(x)\mathrm{d}x \approx \sum_{i=1}^n w_i f(x_i),$$

where the weights w_i and nodes x_i are obtained such that the approximation is exact for a set of $2n$ different functions [83].

To evaluate the four dimensional integrals in Table 4.1, we use the following approximation:

$$\int_0^1 \int_0^1 \int_0^1 \int_0^1 f(x_1,x_2,x_3,x_4)\mathrm{d}x_1 \mathrm{d}x_2 \mathrm{d}x_3 \mathrm{d}x_4 \approx \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{i_3=1}^n \sum_{i_4=1}^n w_{i_1} w_{i_2} w_{i_3} w_{i_4} f(x_{i_1},x_{i_2},x_{i_3},x_{i_4}),$$

---

[7]Note that domains of integrals in Table 4.1 are all $[0,1]$.
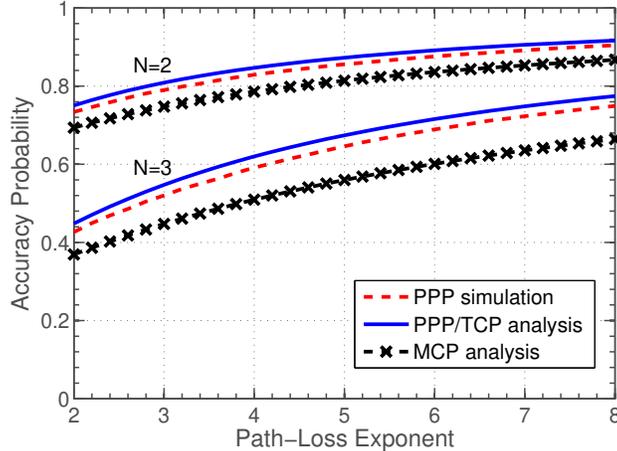
Figure 4.2: Accuracy probability as a function of path-loss exponent for Rayleigh fading.

where 30-point ($n = 30$) Gaussian quadrature rule is employed. The values of the weights w_i and nodes x_i are provided in [84][Table 3].

4.8.2 Simulation Parameters

We consider $\lambda = 0.0005$, $R = 20$, $\sigma^2 = 25$, and $\Omega = 1$. Note that when the numerical results match the simulation results, the numerical results are presented. As we have mentioned in Section 4.2, (4.1) is an approximation for the PDF of the desired link distance in the typical Voronoi cell. Therefore, for the PPP model, we plot both numerical and simulation results. Moreover, as we have mentioned in the previous subsection, four or higher dimensional integrals are approximated using generalized Gaussian quadrature method. Hence, for $N = 3$ in Nakagami-m fading, simulation and analytical results for all PPP, MCP, and TCP models are provided.

4.8.3 Results and Discussions

Impact of Path-Loss Exponent

In Fig. 4.2 and Fig. 4.3, the accuracy probability for Rayleigh and Nakagami-m fading is illustrated as a function of path-loss exponent α. The analytical results are provided in Table 4.1 for two and three users. According to Fig. 4.2, for Rayleigh fading with $\alpha = 4$, ranking users based on their distances for 2-UE NOMA is accurate with probability 0.84 for PPP and TCP. For MCP, ranking users based on their distances is valid with probability 0.79. Therefore, for $N = 2$, ordering users based on their distances instead of instantaneous signal powers seems reasonable. However, for $N = 3$, accuracy probability decreases significantly. When $\alpha = 4$, the accuracy probability is about 0.61 for TCP and PPP and is 0.51 for MCP.

(a) PPP and TCP.　　　　　　　　　　　　(b) MCP.
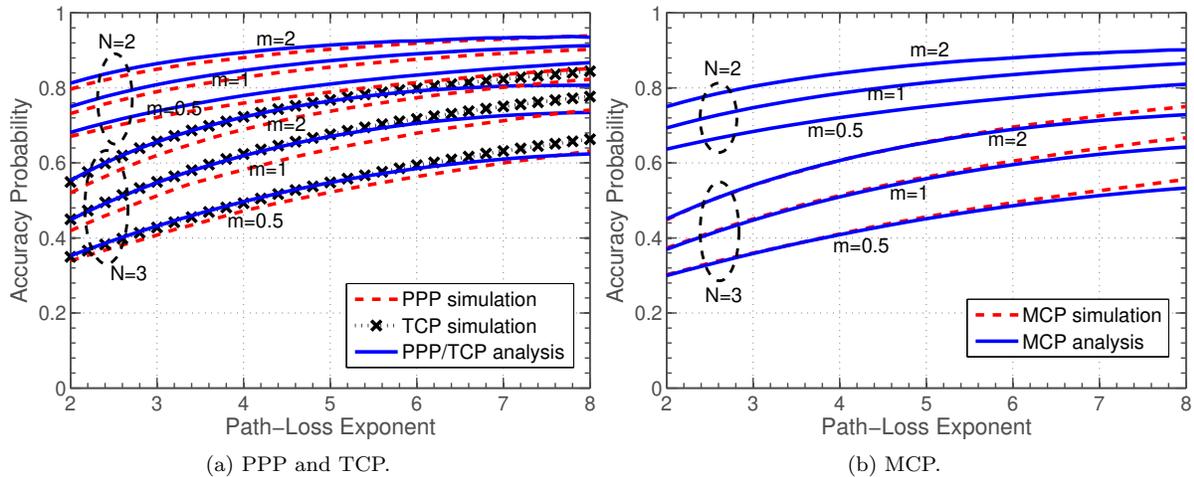
Figure 4.3: Accuracy probability as a function of path-loss exponent for Nakagami-m fading with $m = 0.5, 1, 2$.
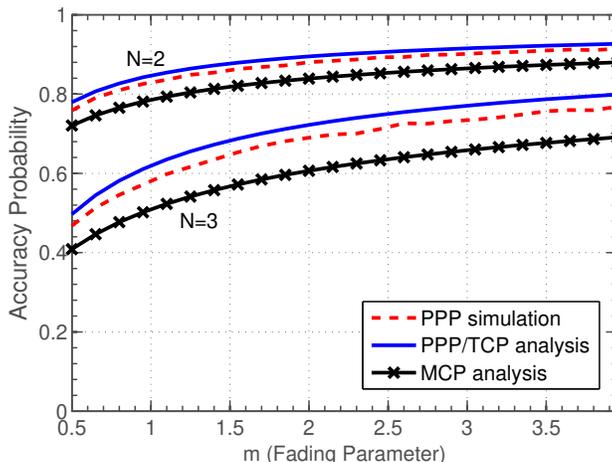
Figure 4.4: Accuracy probability as a function of fading parameter m for $\alpha = 4$.

In Fig. 4.3, for Nakagami-m fading, the accuracy probability is illustrated for different values of m. For $N = 3$, we use the Gaussian quadrature method to numerically evaluate the four dimensional integrals in Table 4.1. Note that, for $N = 3$, the difference between simulation results and analysis for TCP and MCP in Fig. 4.3(a) and Fig. 4.3(b) is due to the Gaussian quadrature method. In summary, we can observe that distance-based ranking yields more accurate coverage probability results for higher values of α, m, and less number of users in a NOMA cluster.

Impact of Fading Parameter m

When $\alpha = 4$, for Nakagami-m fading, in Fig. 4.4, the accuracy probability is shown as a function of m. As we can see, the accuracy probability is an increasing function of m. Therefore, in scenarios with better fading
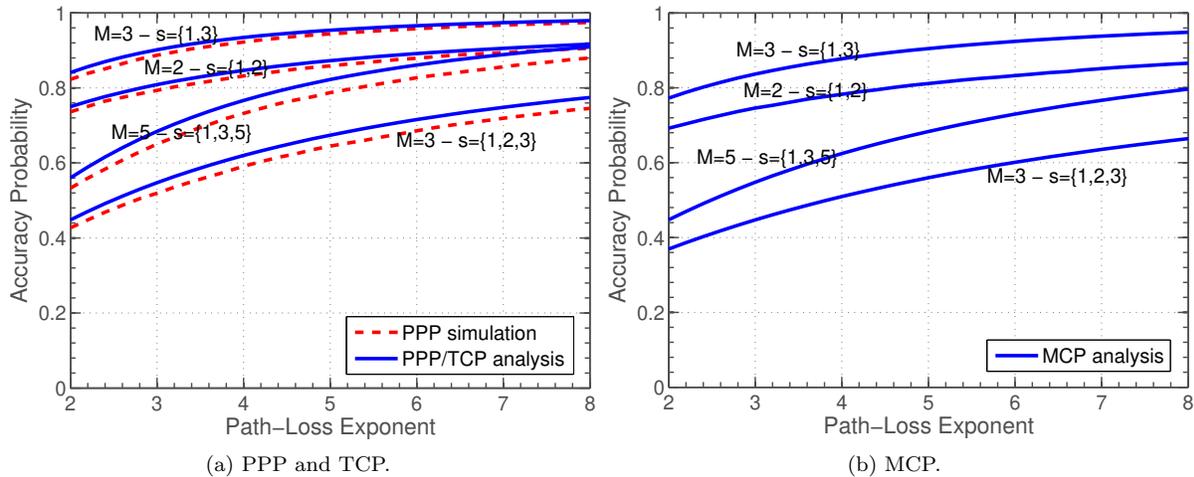
Figure 4.5: Accuracy probability as a function of path-loss exponent and selecting different users for NOMA transmission with Rayleigh fading. M denotes the total number of users out of which N users are selected to form a NOMA cluster. Set s contains ranks of the selected users for a NOMA cluster.

conditions, the distance-based approximation is reasonable. This result is also intuitive because when fading conditions improve, the impact of fading on the channel power is not significant and the distance-based path-loss is dominant. As such, the distance-based approximation is reasonable.

Impact of Distance-Based User Selection

The accuracy probability is shown in Fig. 4.5 for 2-UE and 3-UE NOMA clusters. For instance, when two users are randomly selected, the accuracy probability for PPP with $\alpha = 4$ is about 0.84. However, if we select three users randomly, and then choose the nearest and farthest users to form a NOMA cluster, the accuracy probability will be 0.92. With more associated users with the serving BS, selection of the nearest and farthest users provide higher degree of distinctness among users. According to Fig. 4.5, with increasing channel distinctness, the accuracy probability increases significantly.

4.9 Conclusion

Most of the existing state-of-the-art analyzed NOMA performance assuming that ranking users in each NOMA cluster based on their distances, instead of the complete CSI, is a valid approximation. This approximation affects the coverage probability analysis in the uplink as well as in the downlink. This chapter has verified this assumption for Rayleigh and Nakagami-m fading channels and a variety of users' spatial location distributions such as PPP, MCP, and TCP. Specifically, the accuracy probability, which is the probability that the distance-based ranking matches ranking based on the instantaneous signal power, has

been defined and derived. The results show that the accuracy probability is increasing with respect to the path-loss exponent while it does not depend on the BS intensity in the PPP model, cluster radius in the MCP model, and scattering variance in the TCP model. Effect of user pairing on the accuracy probability has also been investigated, and it has been shown that with distinct user pairing the accuracy probability increases significantly, compared to the random user selection.

Chapter 5

Stochastic Geometry Analysis of Sojourn Time in Multi-Tier Cellular Networks

Impact of mobility will be increasingly important in future generation wireless services and the related challenges will need to be addressed. Sojourn time, the time duration that a mobile user stays within a cell, is a mobility-aware parameter that can significantly impact the performance of mobile users and it can also be exploited to improve resource allocation and mobility management methods in the network. In this chapter, we derive the distribution and mean of the sojourn time in multi-tier cellular networks, where spatial distribution of base stations (BSs) in each tier follows an independent homogeneous Poisson point process (PPP). To obtain the sojourn time distribution in multi-tier cellular networks with maximum biased averaged received power association, as the first step, we derive the area of contact, based on which we then derive the linear contact distribution function and chord length distribution of each tier. We also study the relation between mean sojourn time and other mobility-related performance metrics. We show that the mean sojourn time is inversely proportional to the handoff rate, and the complementary cumulative distribution function (CCDF) of sojourn time is bounded from above by the complement of the handoff probability. Moreover, we study the impact of user velocity and network parameters on the sojourn time.

5.1 Introduction

5.1.1 Background and Related Work

The next generations of cellular wireless networks are expected to support communications for highly mobile users and devices [23] with applications in new vertical sectors such as railway, unmanned aerial vehicle (UAV), and autonomous car. Therefore, addressing the mobility related challenges is necessary for the development of the next generation cellular networks. Impact of user/device mobility on its performance in cellular networks can be measured through mobility-aware performance metrics such as handoff rate, handoff probability, and sojourn time [23]. Sojourn time (or dwell time), time duration that a mobile user stays within a cell, is a key network parameter which allows studying other important network parameters such as channel occupancy time, new call and handoff call dropping probabilities [25]. Therefore, it is imperative to incorporate the sojourn time distribution in resource allocation and mobility management for improving the network performance. In general, modeling and analysis of mobility-related parameters and performances is however challenging in multi-tier (or heterogeneous) cellular networks (e.g. a two-tier macrocell-small cell network) since it needs to consider different aspects such as how to model the distributions of base stations (BSs) at the different tiers, how to model the user mobility and traffic at the different tiers, and how to model the radio access network performance at the different tiers [85].

In this above context, [29] derived the sojourn time distribution for the hexagonal (deterministic) cellular networks and Poisson (random) cellular networks, where the BSs are distributed according to a homogeneous Poisson point process (PPP). In [86], mean sojourn time of two-tier cellular networks was approximately derived, where the coverage areas of macro cells and small cells have regular shapes (circles) and within each macro cell multiple small cells are irregularly deployed. [27, 28] derived the mean sojourn time in small cells of two-tier cellular networks. The BSs of each tier are distributed following an independent homogeneous PPP. [26] also derived the mean sojourn time in two-tier cellular networks. However, it was assumed that a handoff occurs only when the mobile user crosses the boundary of a macro cell. Therefore, the mean sojourn time in [26] is similar to that in a single-tier network as in [29].

Moreover, the handoff rate, i.e. the expected number of handoffs in unit time, was derived in [29] for single-tier Poisson cellular networks and in [87] for multi-tier Poisson cellular networks. The handoff probability, i.e. the probability that the mobile user handoffs to a new BS at the end of a movement period, was also studied in [88] and [89] for single-tier and multi-tier Poisson networks, respectively. To derive the mean sojourn time (or distribution of the sojourn time), [29, 28, 26] used the *chord length distribution* (or *linear contact distribution function*) of Poisson Voronoi cells. However, in multi-tier networks with different transmission power and bias factor for each tier, we need the chord length distribution (or linear contact

59

distribution function) of weighted Poisson Voronoi cells which is not available in the literature. For single-tier networks, [23] used the handoff probability to derive the distribution of sojourn time in the cell where connection is initiated. In single-tier networks, since the Voronoi cells are convex, we can directly use the handoff probability to derive the distribution of the sojourn time. However, in multi-tier networks, cells may not be convex. Therefore, the analytical method in [23] cannot be used for multi-tier networks.

A handoff is considered to be unnecessary when the dwell time of the mobile user in the new cell after the handoff is less than a predefined threshold. In [90, 91], handoff skipping schemes are employed to avoid unnecessary handoffs. Moreover, important system parameters such as channel occupancy time, new call and handoff call dropping probabilities depend on the sojourn time [25]. Therefore, sojourn time is fundamental for analysis and design of the mobile cellular networks. In [90, 25, 92, 93, 94], different distributions such as exponential, Erlang, gamma, Pareto, and Weibull were used for modeling the sojourn time distribution. Due to the principal role of the sojourn time in mobility management and resource allocation, in this chapter, *we derive the sojourn time distribution in multi-tier cellular networks.*

5.1.2 Contributions

To analyze the sojourn time distribution in multi-tier scenarios with PPP distributed BSs, the existing works either assume that the mobile user is always associated to only one of the tiers, or only focus on the small tier (in two-tier scenarios). For both the cases, the results are no different from the single-tier scenarios. In single-tier networks with maximum averaged received power association (nearest BS association), (Voronoi) cells are convex; however, in multi-tier networks with maximum biased averaged received power association, cells may not be convex depending on the transmission power and bias factor of each tier. Therefore, analysis of sojourn time of multi-tier cellular networks is more complicated compared to the single-tier networks. In this regard, the contributions of this chapter can be summarized as follows:

- We derive the distribution and mean of the sojourn time for multi-tier cellular networks. We show that the mean sojourn time is inversely proportional to speed. We also study the impact of network parameters on the sojourn time.

- To obtain the analytical results, we derive the linear contact distribution and chord length distribution of each tier.

- We show that the mean sojourn time is inversely proportional to handoff rate. Also, using handoff rate and sojourn time, we calculate the ping-pong rate (i.e. rate of unnecessary handoffs) for each tier.

- We show that the complement of the handoff probability provides an upper bound for the comple-

mentary cumulative distribution function (CCDF) of the sojourn time. We also discuss the scenarios where the CCDF of the sojourn time is equal to the complement of the handoff probability.

- We show that the mean sojourn time for random mobility models such as random walk converges to the derived results as the user velocity, flight length, or flight time increases.

The rest of this chapter is organized as follows: In Section 5.2, the system model is presented. In Section 5.3, we state the methodology for deriving the analytical results. In Sections 5.4 and 5.5, we obtain the main results related to the distribution and mean of the sojourn time and also discuss the effects of network parameters. Numerical and simulation results are provided in Section 5.6. Finally, in Section 5.7, we conclude the chapter.

5.2 System Model and Notations

Consider a K-tier heterogeneous cellular network with K classes of BSs and let $\mathcal{K} = \{1, 2, ..., K\}$. The spatial distribution of BSs of k-th tier, $k \in \{1, 2, ..., K\}$, follows an independent homogeneous PPP Φ_k of intensity λ_k. Different tiers of BSs transmit at different power levels. P_k denotes the transmission power of the k-th tier BSs.

Consider a typical mobile user which moves in a straight line with a constant speed v. Due to the stationarity of the homogeneous PPP, i.e. its distribution is invariant under translation [41], we can assume that the typical mobile user is located at the origin o at time 0. Since homogeneous PPP is isotropic, i.e. its distribution is invariant under rotation with respect to the origin [41], we can also assume that the typical mobile user moves along the positive x-axis. Therefore, at time t, the typical mobile user is located at $\mathrm{x}(t) = (vt, 0)$.

The mobile user is always associated to the BS which provides the maximum biased averaged received power. Let us denote the serving BS at time t by $BS(t)$. Therefore,

$$BS(t) = \arg \max_{x \in \Phi_k, \forall k \in \mathcal{K}} B_k P_k \|\mathrm{x}(t) - x\|^{-\alpha}, \tag{5.1}$$

where B_k is the cell range expansion bias factor for tier-k, and α is the path-loss exponent. Let us denote the distance between $BS(0)$ and the mobile user at x(t) by $r_0(t)$, i.e. $r_0(t) = \|BS(0) - \mathrm{x}(t)\|$. Given that at time t the mobile user is associated to a tier-k BS at distance $r(t)$, from (5.1), we have $\Phi_j \left(\mathcal{B} \left(\mathrm{x}(t), \frac{r(t)}{\beta_{kj}} \right) \right) = 0$, $\forall j \in \mathcal{K}$, where $\beta_{kj} = \left(\frac{B_k P_k}{B_j P_j} \right)^{1/\alpha}$, $\mathcal{B}(\mathrm{x}, r)$ denotes a ball with radius r centered at x, and $\Phi_j(A)$ is the number of tier-j BSs in set $A \subset \mathbb{R}^2$. For simplicity, we define $r_0 = r_0(0)$.

A summary of the major notations is provided in Table 5.1.

Table 5.1: Summary of Notations

Notation	Description
Φ_k, λ_k	PPP of tier-k BSs, intensity of Φ_k
P_k	Transmit power of the k-th tier BSs
x(t)	Location of the mobile user at time t
v	Speed of the mobile user
$BS(t)$	Serving BS of the mobile user at time t
B_k	Cell range expansion (bias) factor for tier k
α	Path-loss exponent
$r_0(t)$, r_0	Distance between the initially serving BS and the mobile user at time t, $r_0(0)$
β_{kj}	$\left(\frac{B_k P_k}{B_j P_j}\right)^{1/\alpha}$
$\mathcal{B}(\mathrm{x}, r)$	Ball with radius r centred at x
$\tilde{S}$	Sojourn time in the cell where connection is initiated
S	Sojourn time
H_k	Handoff rate from (to) a tier-k cell to (from) any other cell in the network
H	Handoff rate

5.3 Methodology of Analysis of Sojourn Time in Multi-Tier Cellular Networks

The sojourn time S is the duration that the mobile user stays within a particular serving cell before it is handed over to another cell [95]. Analysis of sojourn time in multi-tier cellular networks consists of the following four steps:

- **Step 1**: Deriving the conditional distribution of the sojourn time in the cell where connection is initiated, given that the mobile user is initially associated to a tier-k BS.

- **Step 2**: Deriving the linear contact distribution function, given that the mobile user is in a tier-k cell at time 0.

- **Step 3**: Obtaining the chord length distribution for tier k using linear contact distribution function.

- **Step 4**: Deriving the distribution of the sojourn time S for tier k.

(a) Single-tier.   (b) Two-tier.
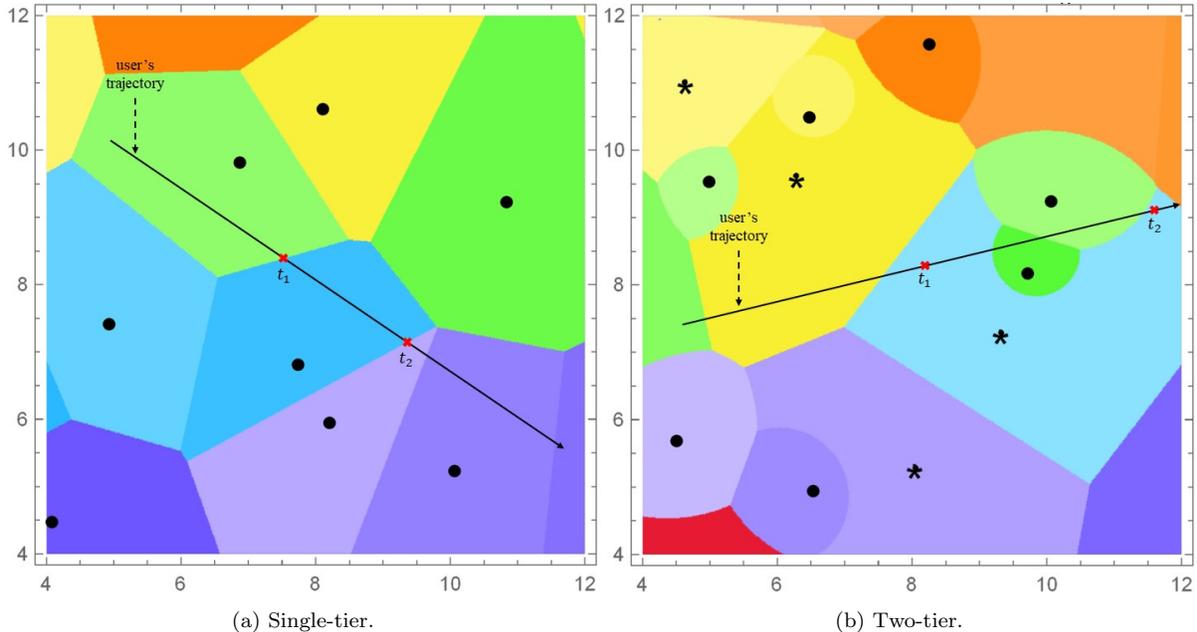
Figure 5.1: (Weighted) Voronoi cells in a single-tier and a two-tier cellular network: (a) $\lambda = 0.15$, (b) $\lambda_1 = 0.05$, $\lambda_2 = 0.1$, and $\beta_{12} = 2$. In (a), BSs are represented by black circles and user's trajectory is shown with a black solid line. Since the cells are convex, there is no handoff in the interval $[t_1, t_2]$ when the same BS serves the user at time instants t_1 and t_2, i.e. the same BS serves the user at any time between t_1 and t_2 if and only if it serves the user at t_1 and t_2. In (b), we show tier-one BSs by black asterisks and tier-two BSs by black circles. As can be seen, although the same BS serves the user at t_1 and t_2, a different BS may serve the user at time $t \in (t_1, t_2)$. Therefore, being served by the same BS at t_1 and t_2 is necessary but not sufficient for occurrence of no handoff in the interval $[t_1, t_2]$. This stems from the fact that, in a multi-tier network, the weighted Voronoi cells are not convex when the BPs (product of bias factor and transmission power) are different for different tiers.

5.3.1   Step 1 of Analysis

First we focus on the distribution of the sojourn time in the cell where connection is initiated $\tilde{S}$. Specifically, we derive the CCDF of $\tilde{S}$, i.e.

$$\bar{F}_{\tilde{S}}(T) = \mathbb{P}(\tilde{S} > T) = \mathbb{P}\left(\text{no handoff occurs in the interval } [0, T]\right)$$

$$= \mathbb{P}\left(BS(t) = BS(0), \forall t \in (0, T]\right). \tag{5.2}$$

In single-tier cellular networks, Voronoi cells are convex [96] (as shown in Fig. 5.1(a)). A set C is convex if the line segment between any two points in C lies in C [97]. Thus, in single-tier cellular networks, when the mobile user is connected to the same BS at time 0 and T, i.e. when $BS(T) = BS(0)$, the serving BS at any time between 0 and T is also $BS(0)$, i.e. $BS(t) = BS(0), \forall t \in (0, T)$. Hence, for single-tier cellular
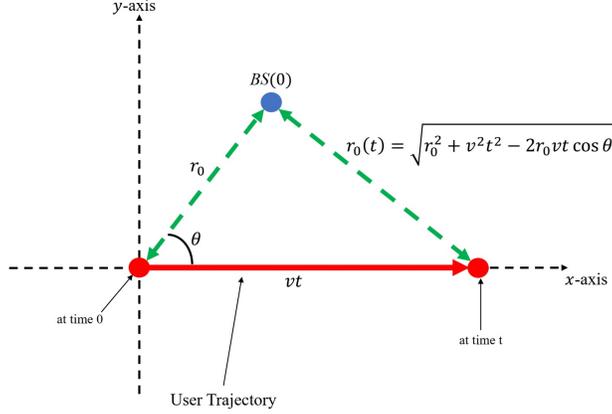
Figure 5.2: System model.

networks, (5.2) can be simplified as

$$\bar{F}_{\tilde{S}}(T) = \mathbb{P}\left(BS(T) = BS(0)\right)^1. \tag{5.3}$$

However, for multi-tier cellular networks, Voronoi cells may not be convex depending on the values of β_{kj}, $k, j \in \mathcal{K}$ (as shown in Fig. 5.1(b)). Therefore, even when $BS(T) = BS(0)$, there may exist a time t between 0 and T for which $BS(t) \neq BS(0)$. To derive the CCDF of $\tilde{S}$ for multi-tier cellular networks, we must use (5.2), which makes the analysis of sojourn time in multi-tier cellular networks more complicated compared to the single-tier networks. Actually, single-tier scenario can be considered as a special case of multi-tier scenarios. Moreover, note that, (5.3) provides an upper bound for (5.2).

Given that the mobile user is initially connected to a tier-k BS, the CCDF of $\tilde{S}$ can be obtained by

$$\begin{aligned}\bar{F}_{\tilde{S}}(T \mid \text{tier} = k) &= \mathbb{P}\left(BS(t) = BS(0), \forall t \in (0, T] \mid \text{tier} = k\right) \\ &= \frac{1}{\pi} \int_0^\infty \int_0^\pi \mathbb{P}\left(BS(t) = BS(0), \forall t \in (0, T] \mid r_0, \theta, \text{tier} = k\right) f_R(r_0 \mid \text{tier} = k) \mathrm{d}\theta \mathrm{d}r_0,\end{aligned} \tag{5.4}$$

where θ is the angle between the serving BS at time 0 and direction of the movement (as shown in Fig. 5.2). θ is uniformly distributed in $[0, \pi]$. $f_R(r_0 \mid \text{tier} = k)$ is the probability density function (PDF) of the serving link distance at time 0, given that $BS(0)$ belongs to tier-k. According to [98],

$$f_R(r_0 \mid \text{tier} = k) = \frac{1}{\mathbb{P}(\text{tier} = k)} 2\lambda_k \pi r_0 \exp\left\{-\sum_{j \in \mathcal{K}} \lambda_j \pi \beta_{jk}^2 r_0^2\right\}, \tag{5.5}$$

---

[1] $\mathbb{P}\left(BS(T) \neq BS(0)\right)$ is the probability that the mobile user is handed off to a new BS after the movement period of T, and it is called handoff probability in the literature. According to (5.3), in single-tier networks, the CCDF of the sojourn time is equal to the complement of the handoff probability.

where $\mathbb{P}(\text{tier} = k)$ is the probability that $BS(0)$ belongs to tier-k which is given by [98]:

$$\mathbb{P}(\text{tier} = k) = \frac{\lambda_k}{\sum_{j \in \mathcal{K}} \lambda_j \beta_{jk}^2}. \tag{5.6}$$

Using the association strategy (5.1), we get

$$\mathbb{P}\left(BS(t) = BS(0), \forall t \in (0, T] \mid r_0, \theta, \text{tier} = k\right)$$

$$= \mathbb{P}\left(\bigcap_{j \in \mathcal{K}} \Phi_j \left(\mathcal{B}\left(\mathbf{x}(t), \frac{r_0(t)}{\beta_{kj}}\right) \setminus \mathcal{B}\left(0, \frac{r_0}{\beta_{kj}}\right)\right) = 0, \forall t \in (0, T] \mid r_0, \theta, \text{tier} = k\right), \tag{5.7}$$

where $\mathcal{B}\left(0, \frac{r_0}{\beta_{kj}}\right)$ is excluded since we know there is no tier j BS closer than $\frac{r_0}{\beta_{kj}}$ to the typical mobile user at time 0. Let us define

$$\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj}) = \left\{\bigcup_t \mathcal{B}\left(\mathbf{x}(t), \frac{r_0(t)}{\beta_{kj}}\right) \mid t \in [0, T], r_0(t) = \sqrt{r_0^2 + v^2 t^2 - 2r_0 vt \cos\theta}\right\}. \tag{5.8}$$

Using $\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj})$, we can write

$$\mathbb{P}\left(BS(t) = BS(0), \forall t \in (0, T] \mid r_0, \theta, \text{tier} = k\right)$$

$$= \mathbb{P}\left(\bigcap_{j \in \mathcal{K}} \Phi_j \left(\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj}) \setminus \mathcal{B}\left(0, \frac{r_0}{\beta_{kj}}\right)\right) = 0 \mid r_0, \theta, \text{tier} = k\right)$$

$$\overset{(a)}{=} \prod_{j \in \mathcal{K}} \mathbb{P}\left(\Phi_j \left(\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj}) \setminus \mathcal{B}\left(0, \frac{r_0}{\beta_{kj}}\right)\right) = 0 \mid r_0, \theta, \text{tier} = k\right)$$

$$\overset{(b)}{=} \prod_{j \in \mathcal{K}} \exp\left\{-\lambda_j \left|\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj}) \setminus \mathcal{B}\left(0, \frac{r_0}{\beta_{kj}}\right)\right|\right\}, \tag{5.9}$$

where $|A|$ denotes the area of A, (a) follows from the independence of different tiers' point processes, and (b) is obtained by using the void probability of PPP. In Fig. 5.3, $\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj})$ is illustrated for three different cases: a) $\beta_{kj} < 1$, b) $\beta_{kj} = 1$, and c) $\beta_{kj} > 1$. To derive the distribution of $\tilde{S}$, we need to calculate the area of $\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj})$ for all three cases. Further discussion about $\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj})$ is provided in the next section.

5.3.2 Step 2 of Analysis

Given that, at time 0, the mobile user is associated to a tier-k BS, the origin is almost surely contained in the interior of a tier-k Voronoi cell. In this chapter, we define *linear contact distribution function* as the probability that a line segment ℓ containing the origin with length r and random orientation crosses

(a) $\beta_{kj} = 0.8$.   (b) $\beta_{kj} = 1$.   (c) $\beta_{kj} = 1.2$.
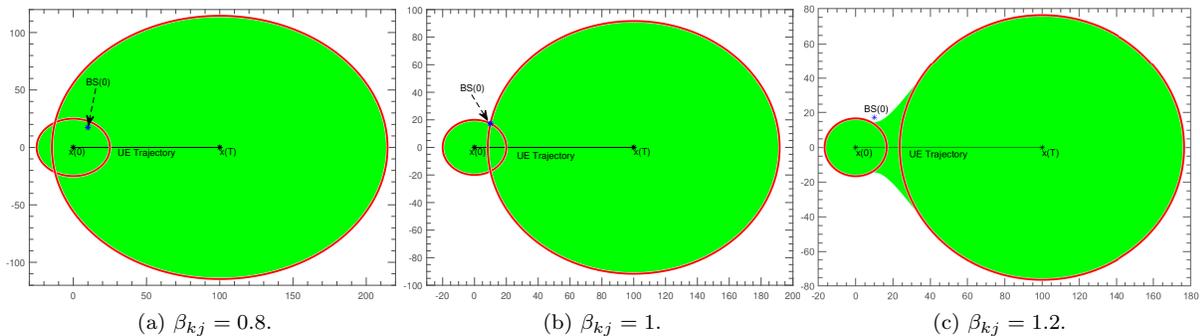
Figure 5.3: $\mathcal{A}_{kj}(20, \pi/3, 5, 20, \beta_{kj})$. (a) $\beta_{kj} < 1$, (b) $\beta_{kj} = 1$, and (c) $\beta_{kj} > 1$. Red circles correspond to $\mathcal{B}\left(0, \frac{r_0}{\beta_{kj}}\right)$ and $\mathcal{B}\left(\mathrm{x}(T), \frac{r_0(T)}{\beta_{kj}}\right)$.

the cell boundaries. Therefore, given origin o is inside a tier-k cell, linear contact distribution function $H_\ell(z \mid \text{tier} = k)$ is equal to the probability that intersection of user's trajectory with length z and the cell boundaries is nonempty. Using the conditional CCDF of $\tilde{S}$, we can derive the linear contact distribution function as

$$H_\ell(z \mid \text{tier} = k) = 1 - \mathbb{P}\left(\tilde{S} > \frac{z}{v} \mid \text{tier} = k\right) = 1 - \bar{F}_{\tilde{S}}\left(\frac{z}{v} \mid \text{tier} = k\right). \tag{5.10}$$

5.3.3 Step 3 of Analysis

So far, we have considered the sojourn time in the cell where connection is initiated ($\tilde{S}$). Distribution of the sojourn time (S), for tier-k, can be obtained using the chord length distribution. Due to the stationarity of our model, chord length distribution for tier-k, denoted by $F_L(z \mid \text{tier} = k)$, can be computed as follows [42]:

$$F_L(z \mid \text{tier} = k) = 1 - \mathbb{E}[L \mid \text{tier} = k] \frac{\mathrm{d}}{\mathrm{d}z} H_\ell(z \mid \text{tier} = k), \tag{5.11}$$

where $\mathbb{E}[L \mid \text{tier} = k]$ is the mean length of the chords lying in tier-k cells, and is obtained by [99]

$$\mathbb{E}[L \mid \text{tier} = k] = \lim_{z \to 0} \frac{z}{H_\ell(z \mid \text{tier} = k)} \tag{5.12}$$

5.3.4 Step 4 of Analysis

Finally, we can characterize the sojourn time distribution for tier-k using the results from previous step. In particular, the mean and CCDF of the sojourn time in tier-k are

$$\mathbb{E}[S \mid \text{tier} = k] = \frac{1}{v}\mathbb{E}[L \mid \text{tier} = k], \tag{5.13}$$

$$\bar{F}_S(T \mid \text{tier} = k) = 1 - F_L(vT \mid \text{tier} = k). \tag{5.14}$$

5.4 Derivation of $|\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj})|$

As mentioned in the previous section, the first step of sojourn time analysis requires calculation of area of $\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj})$. In this regard, we consider three cases: I) $\beta_{kj} < 1$, II) $\beta_{kj} > 1$, and III) $\beta_{kj} = 1$.

5.4.1 Case I: $\beta_{kj} < 1$

The following proposition helps us to derive the area of $\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj})$ for this case.

Proposition 5.1. *When* $\beta_{kj} < 1$, $\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj}) = \mathcal{B}\left(\mathbf{x}(0), \frac{r_0}{\beta_{kj}}\right) \cup \mathcal{B}\left(\mathbf{x}(T), \frac{r_0(T)}{\beta_{kj}}\right)$.

Proof. See **Appendix C.1**. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Note that, depending on radii of the two circles, $\frac{r_0}{\beta_{kj}}$ and $\frac{r_0(T)}{\beta_{kj}}$, and the distance between their centres, i.e. vT, three different situations can happen when $\beta_{kj} < 1$:

Situation 1: When $\frac{r_0(T)}{\beta_{kj}} \geq \frac{r_0}{\beta_{kj}} + vT$, we have $\mathcal{B}\left(\mathbf{x}(0), \frac{r_0}{\beta_{kj}}\right) \subset \mathcal{B}\left(\mathbf{x}(T), \frac{r_0(T)}{\beta_{kj}}\right)$, which yields

$$\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj}) = \mathcal{B}\left(\mathbf{x}(T), \frac{r_0(T)}{\beta_{kj}}\right).$$

Situation 2: When $\frac{r_0}{\beta_{kj}} \geq \frac{r_0(T)}{\beta_{kj}} + vT$, we have $\mathcal{B}\left(\mathbf{x}(T), \frac{r_0(T)}{\beta_{kj}}\right) \subset \mathcal{B}\left(\mathbf{x}(0), \frac{r_0}{\beta_{kj}}\right)$, which yields

$$\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj}) = \mathcal{B}\left(\mathbf{x}(0), \frac{r_0}{\beta_{kj}}\right).$$

Situation 3: When $\frac{r_0(T)}{\beta_{kj}} < \frac{r_0}{\beta_{kj}} + vT$ and $\frac{r_0}{\beta_{kj}} < \frac{r_0(T)}{\beta_{kj}} + vT$,

$$\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj}) = \mathcal{B}\left(\mathbf{x}(0), \frac{r_0}{\beta_{kj}}\right) \cup \mathcal{B}\left(\mathbf{x}(T), \frac{r_0(T)}{\beta_{kj}}\right).$$

An example of which is illustrated in Fig. 5.3(a).

(a) $t = 2, 2.2$.   (b) $t = 0, 1, 2, 3, 4, 5, 6, 7, 8$.
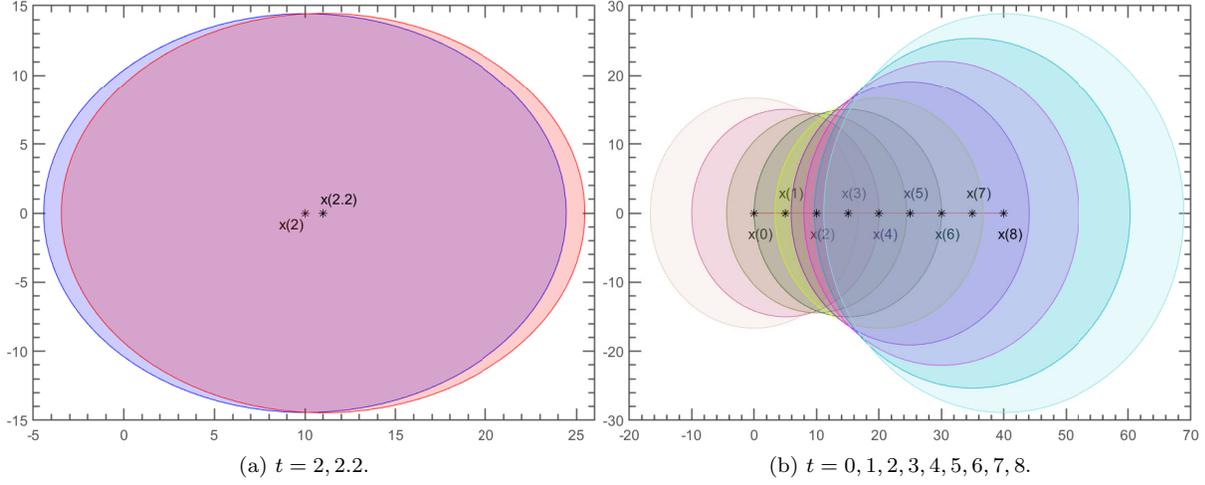
Figure 5.4: $\mathcal{B}\left(\mathrm{x}(t), \frac{r_0(t)}{\beta_{kj}}\right)$ at different time instants. Union of these circles from $t = 0$ till $t = 8$ forms $\mathcal{A}_{kj}(20, \pi/3, 5, 8, 1.2)$.

Using this information, now we can compute $|\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj})|$ when $\beta_{kj} < 1$.

$$
|\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj})| = \begin{cases} \pi \frac{r_0(T)^2}{\beta_{kj}^2}, & \text{if } 2r_0 \frac{\cos\theta + \beta_{kj}}{1 - \beta_{kj}^2} \leq vT \\[2mm] \pi \frac{r_0^2}{\beta_{kj}^2}, & \text{if } vT \leq 2r_0 \frac{\cos\theta - \beta_{kj}}{1 - \beta_{kj}^2} \\[2mm] \pi \frac{r_0^2}{\beta_{kj}^2} + \pi \frac{r_0(T)^2}{\beta_{kj}^2} - V\left(\frac{r_0}{\beta_{kj}}, \frac{r_0(T)}{\beta_{kj}}, vT\right), & \text{if } 2r_0 \frac{\cos\theta - \beta_{kj}}{1 - \beta_{kj}^2} < vT < 2r_0 \frac{\cos\theta + \beta_{kj}}{1 - \beta_{kj}^2} \end{cases} \tag{5.15}
$$

where $V\left(\frac{r_0}{\beta_{kj}}, \frac{r_0(T)}{\beta_{kj}}, vT\right)$ is the area of intersection of two circles with radii $\frac{r_0}{\beta_{kj}}$ and $\frac{r_0(T)}{\beta_{kj}}$ whose centers are separated by vT, i.e. $V\left(\frac{r_0}{\beta_{kj}}, \frac{r_0(T)}{\beta_{kj}}, vT\right) =$

$$
\frac{r_0^2}{\beta_{kj}^2} \arccos\left(\frac{r_0^2 + \beta_{kj}^2 v^2 T^2 - r_0(T)^2}{2\beta_{kj} r_0 vT}\right) + \frac{r_0(T)^2}{\beta_{kj}^2} \arccos\left(\frac{r_0(T)^2 + \beta_{kj}^2 v^2 T^2 - r_0^2}{2\beta_{kj} r_0(T) vT}\right)
$$
$$
- \frac{1}{2}\sqrt{\left(\frac{r_0}{\beta_{kj}} + \frac{r_0(T)}{\beta_{kj}} + vT\right)\left(\frac{r_0}{\beta_{kj}} + \frac{r_0(T)}{\beta_{kj}} - vT\right)\left(\frac{r_0}{\beta_{kj}} - \frac{r_0(T)}{\beta_{kj}} + vT\right)\left(-\frac{r_0}{\beta_{kj}} + \frac{r_0(T)}{\beta_{kj}} + vT\right)}. \tag{5.16}
$$

5.4.2 Case II: $B_{kj} > 1$

For this case, to derive the area of $\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj})$, first we study the intersection of $\mathcal{B}\left(\mathrm{x}(t), \frac{r_0(t)}{\beta_{kj}}\right)$ and $\mathcal{B}\left(\mathrm{x}(t + \mathrm{d}t), \frac{r_0(t + \mathrm{d}t)}{\beta_{kj}}\right)$ as $\mathrm{d}t \to 0$ (Fig. 5.4(a)). From triangle equations, we have

$$
r_0(t + \mathrm{d}t)^2 = r_0^2 + v^2(t + \mathrm{d}t)^2 - 2r_0 v(t + \mathrm{d}t)\cos\theta = r_0(t)^2 + v^2 \mathrm{d}t^2 + 2v\mathrm{d}t(vt - r_0\cos\theta). \tag{5.17}
$$

68

Since $|vt - r_0 \cos\theta| \leq r_0(t)$,

$$r_0(t)^2 + v^2\mathrm{d}t^2 - 2r_0(t)v\mathrm{d}t \leq r_0(t+\mathrm{d}t)^2 \leq r_0(t)^2 + v^2\mathrm{d}t^2 + 2r_0(t)v\mathrm{d}t. \tag{5.18}$$

Dividing $r_0(t) - v\mathrm{d}t \leq r_0(t+\mathrm{d}t) \leq r_0(t) + v\mathrm{d}t$ by β_{kj} yields

$$\frac{r_0(t)}{\beta_{kj}} - v\mathrm{d}t \overset{(a)}{\leq} \frac{r_0(t)}{\beta_{kj}} - \frac{v\mathrm{d}t}{\beta_{kj}} \leq \frac{r_0(t+\mathrm{d}t)}{\beta_{kj}} \leq \frac{r_0(t)}{\beta_{kj}} + \frac{v\mathrm{d}t}{\beta_{kj}} \overset{(b)}{\leq} \frac{r_0(t)}{\beta_{kj}} + v\mathrm{d}t,$$

where (a) and (b) are obtained using $\beta_{kj} > 1$. Therefore, as $\mathrm{d}t \to 0$, $\mathcal{B}\left(\mathrm{x}(t), \frac{r_0(t)}{\beta_{kj}}\right)$ and $\mathcal{B}\left(\mathrm{x}(t+\mathrm{d}t), \frac{r_0(t+\mathrm{d}t)}{\beta_{kj}}\right)$ partially overlap (boundaries of $\mathcal{B}\left(\mathrm{x}(t), \frac{r_0(t)}{\beta_{kj}}\right)$ and $\mathcal{B}\left(\mathrm{x}(t+\mathrm{d}t), \frac{r_0(t+\mathrm{d}t)}{\beta_{kj}}\right)$ intersect at two points), and we have

$$\left| \mathcal{B}\left(\mathrm{x}(t), \frac{r_0(t)}{\beta_{kj}}\right) \setminus \mathcal{B}\left(\mathrm{x}(t+\mathrm{d}t), \frac{r_0(t+\mathrm{d}t)}{\beta_{kj}}\right) \right| = \pi \frac{r_0(t)^2}{\beta_{kj}^2} - V\left(\frac{r_0(t)}{\beta_{kj}}, \frac{r_0(t+\mathrm{d}t)}{\beta_{kj}}, v\mathrm{d}t\right)$$

$$= \pi \frac{r_0(t)^2}{\beta_{kj}^2} - \arccos\left(\frac{r_0\cos\theta - vt}{\beta_{kj}r_0(t)} + \frac{\beta_{kj}^2 - 1}{2\beta_{kj}}\frac{v\mathrm{d}t}{r_0(t)}\right)\frac{r_0(t)^2}{\beta_{kj}^2}$$

$$- \arccos\left(\frac{vt - r_0\cos\theta}{\beta_{kj}r_0(t+\mathrm{d}t)} + \frac{\beta_{kj}^2 + 1}{2\beta_{kj}}\frac{v\mathrm{d}t}{r_0(t+\mathrm{d}t)}\right)\frac{r_0(t+\mathrm{d}t)^2}{\beta_{kj}^2}$$

$$+ \frac{1}{2}\sqrt{\frac{2v\mathrm{d}t}{\beta_{kj}}\left(r_0(t) - \frac{vt - r_0\cos\theta}{\beta_{kj}}\right) + v^2\mathrm{d}t^2\left(1 - \frac{1}{\beta_{kj}^2}\right)}$$

$$\times \sqrt{\frac{2v\mathrm{d}t}{\beta_{kj}}\left(r_0(t) + \frac{vt - r_0\cos\theta}{\beta_{kj}}\right) - v^2\mathrm{d}t^2\left(1 - \frac{1}{\beta_{kj}^2}\right)}$$

$$= \frac{2v}{\beta_{kj}^2}\left[\sqrt{\beta_{kj}^2 r_0(t)^2 - (vt - r_0\cos\theta)^2} - \arccos\left(\frac{vt - r_0\cos\theta}{\beta_{kj}r_0(t)}\right)(vt - r_0\cos\theta)\right]\mathrm{d}t + O(\mathrm{d}t^2), \tag{5.19}$$

where the last result is proved in **Appendix C.2**.

Moreover, we can derive the intersection points of these two circles from their equations in Cartesian coordinate system, i.e.

$$\mathcal{B}\left(\mathrm{x}(t), \frac{r_0(t)}{\beta_{kj}}\right) : [x - vt]^2 + y^2 = \frac{r_0(t)^2}{\beta_{kj}^2},$$

$$\mathcal{B}\left(\mathrm{x}(t+\mathrm{d}t), \frac{r_0(t+\mathrm{d}t)}{\beta_{kj}}\right) : [x - v(t+\mathrm{d}t)]^2 + y^2 = \frac{r_0(t+\mathrm{d}t)^2}{\beta_{kj}^2}.$$

Combining these equations and solving for x results in

$$x = v\left(t + \frac{\mathrm{d}t}{2}\right)\left(1 - \frac{1}{\beta_{kj}^2}\right) + \frac{r_0\cos\theta}{\beta_{kj}^2},$$

which indicates that, for $\beta_{kj} > 1$, the boundaries' intersection points move along the positive x-axis as t increases (Fig. 5.4(b)). Using this result, we can write

$$\left| \bigcup_{i=0}^{n} \mathcal{B}\left(\mathrm{x}(t+i\mathrm{d}t), \frac{r_0(t+i\mathrm{d}t)}{\beta_{kj}} \right) \right| = \pi \frac{r_0(t+n\mathrm{d}t)^2}{\beta_{kj}^2}$$
$$+ \sum_{i=0}^{n-1} \left| \mathcal{B}\left(\mathrm{x}(t+i\mathrm{d}t), \frac{r_0(t+i\mathrm{d}t)}{\beta_{kj}} \right) \setminus \mathcal{B}\left(\mathrm{x}(t+(i+1)\mathrm{d}t), \frac{r_0(t+(i+1)\mathrm{d}t)}{\beta_{kj}} \right) \right|. \tag{5.20}$$

Let the time interval $[0, T]$ be partitioned by points $t_i = i\mathrm{d}t$, $i = 0, ..., \frac{T}{\mathrm{d}t}$. We can calculate the area of $\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj})$ by setting $t = 0$ in (5.20), i.e.

$$|\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj})| = \lim_{\mathrm{d}t \to 0} \left| \bigcup_{i=0}^{\frac{T}{\mathrm{d}t}} \mathcal{B}\left(\mathrm{x}(i\mathrm{d}t), \frac{r_0(i\mathrm{d}t)}{\beta_{kj}} \right) \right|$$

$$= \pi \frac{r_0(T)^2}{\beta_{kj}^2} + \lim_{\mathrm{d}t \to 0} \sum_{i=0}^{\frac{T}{\mathrm{d}t}-1} \left| \mathcal{B}\left(\mathrm{x}(i\mathrm{d}t), \frac{r_0(i\mathrm{d}t)}{\beta_{kj}} \right) \setminus \mathcal{B}\left(\mathrm{x}((i+1)\mathrm{d}t), \frac{r_0((i+1)\mathrm{d}t)}{\beta_{kj}} \right) \right|$$

$$\overset{(a)}{=} \pi \frac{r_0(T)^2}{\beta_{kj}^2} + \lim_{\mathrm{d}t \to 0} \sum_{i=0}^{\frac{T}{\mathrm{d}t}-1}$$

$$\frac{2v}{\beta_{kj}^2} \left[\sqrt{\beta_{kj}^2 r_0(t_i)^2 - (vt_i - r_0 \cos\theta)^2} - \arccos\left(\frac{vt_i - r_0 \cos\theta}{\beta_{kj} r_0(t_i)} \right) (vt_i - r_0 \cos\theta) \right] \mathrm{d}t + O(\mathrm{d}t^2)$$

$$\overset{(b)}{=} \pi \frac{r_0(T)^2}{\beta_{kj}^2} + \frac{2v}{\beta_{kj}^2} \int_0^T \sqrt{\beta_{kj}^2 r_0(t)^2 - (vt - r_0 \cos\theta)^2} - \arccos\left(\frac{vt - r_0 \cos\theta}{\beta_{kj} r_0(t)} \right) (vt - r_0 \cos\theta)\mathrm{d}t, \tag{5.21}$$

where (a) is obtained by using (5.19) and (b) follows from the Riemann integral.

5.4.3 Case III: $\beta_{kj} = 1$

For this case, similar to **Appendix C.1**, we can prove $\mathcal{A}_{kj}(r_0, \theta, v, T, 1) = \mathcal{B}(\mathrm{x}(0), r_0) \cup \mathcal{B}(\mathrm{x}(T), r_0(T))$. Since $|r_0 - vT| \le r_0(T) \le r_0 + vT$, $\mathcal{B}(\mathrm{x}(0), r_0)$ and $\mathcal{B}(\mathrm{x}(T), r_0(T))$ partially overlap. Therefore,

$$|\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj})| = \pi \frac{r_0^2}{\beta_{kj}^2} + \pi \frac{r_0(T)^2}{\beta_{kj}^2} - V\left(\frac{r_0}{\beta_{kj}}, \frac{r_0(T)}{\beta_{kj}}, vT \right), \tag{5.22}$$

where $\beta_{kj} = 1$ and $V\left(\frac{r_0}{\beta_{kj}}, \frac{r_0(T)}{\beta_{kj}}, vT \right)$ is given in (5.16).

5.4.4 Closed-form Expression for $|\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj})|$

Theorem 5.1. *Area of $\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj})$ can be obtained by*

$$|\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj})| = |\mathcal{A}_{kj}(r_0, \theta, vT, 1, \beta_{kj})| =$$

$$\begin{cases} \pi \frac{g(vT,1)^2}{\beta_{kj}^2} + \frac{2vT}{\beta_{kj}^2} \int\limits_0^1 \sqrt{\beta_{kj}^2 g(vT,u)^2 - (vTu - r_0\cos\theta)^2} - \arccos\left(\frac{vTu - r_0\cos\theta}{\beta_{kj}g(vT,u)}\right)(vTu - r_0\cos\theta)\mathrm{d}u, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \text{if } (\beta_{kj} > 1) \\[6pt] \pi \frac{r_0^2}{\beta_{kj}^2} + \pi \frac{g(vT,1)^2}{\beta_{kj}^2} - V\left(\frac{r_0}{\beta_{kj}}, \frac{g(vT,1)}{\beta_{kj}}, vT\right), \\ \qquad\qquad\qquad \text{if } (\beta_{kj} = 1) \ \ or \ \left(\beta_{kj} < 1 \ and \ 2r_0\frac{\cos\theta - \beta_{kj}}{1-\beta_{kj}^2} < vT < 2r_0\frac{\cos\theta + \beta_{kj}}{1-\beta_{kj}^2}\right) \\[6pt] \pi \frac{r_0^2}{\beta_{kj}^2}, \qquad\qquad\qquad\qquad\qquad \text{if } \left(\beta_{kj} < 1 \ and \ vT \le 2r_0\frac{\cos\theta - \beta_{kj}}{1-\beta_{kj}^2}\right) \\[6pt] \pi \frac{g(vT,1)^2}{\beta_{kj}^2}, \qquad\qquad\qquad\qquad \text{if } \left(\beta_{kj} < 1 \ and \ 2r_0\frac{\cos\theta + \beta_{kj}}{1-\beta_{kj}^2} \le vT\right) \end{cases}, \tag{5.23}$$

where $g(vT, u) = r_0(Tu) = \sqrt{r_0^2 + v^2T^2u^2 - 2r_0vTu\cos\theta}$ and $V\left(\frac{r_0}{\beta_{kj}}, \frac{g(vT,1)}{\beta_{kj}}, vT\right)$ is given in (5.16).

Proof. The proof follows from combining (5.15), (5.21), and (5.22). For, $\beta_{kj} > 1$, we have used change of variable $\frac{t}{T} = u$. $\qquad\square$

It is worth mentioning that, according to **Theorem 5.1**, in multi-tier networks, tier-k cells are convex, if $B_k P_k \le B_j P_j$ (or equivalently, $\beta_{kj} \le 1$), $\forall j \in \mathcal{K}$. Therefore, in this case, we can derive the sojourn time distribution of tier-k similar to the single-tier scenario using (5.3) (instead of (5.2)). This is the reason why some works in the literature only focus on the sojourn time in small cells of two-tier networks.

5.5 Main Results on Sojourn Time and Handoff Rate and Effects of Network Parameters

5.5.1 Conditional CCDF and Mean of Sojourn Time

Since $\mathcal{B}\left(0, \frac{r_0}{\beta_{kj}}\right) \subset \mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj})$, we can further simplify (5.9) as

$$\mathbb{P}\left(BS(t) = BS(0), \forall t \in (0, T] \mid r_0, \theta, \text{tier} = k\right) = \exp\left\{-\sum_{j\in\mathcal{K}} \lambda_j \left(|\mathcal{A}_{kj}(r_0, \theta, vT, 1, \beta_{kj})| - \pi\frac{r_0^2}{\beta_{kj}^2}\right)\right\}, \tag{5.24}$$

The CCDF of the sojourn time of a connection in a cell where it is initiated, $\tilde{S}$ can be obtained by substituting (5.5) and (5.24) in (5.4).

$$\bar{F}_{\tilde{S}}(T \mid \text{tier} = k) = \frac{1}{\mathbb{P}(\text{tier} = k)} \int_0^\infty \int_0^\pi 2\lambda_k r_0 \exp\left\{-\sum_{j\in\mathcal{K}} \lambda_j \left|\mathcal{A}_{kj}(r_0, \theta, vT, 1, \beta_{kj})\right|\right\} d\theta dr_0, \tag{5.25}$$

where $\mathbb{P}(\text{tier} = k)$ is given in (5.6), and $|\mathcal{A}_{kj}(r_0, \theta, vT, 1, \beta_{kj})|$ is given in **Theorem 5.1**.

As discussed before, in Step 2, we use (5.25) to derive the linear contact distribution function given that the mobile user is in a tier-k cell at time 0, i.e.

$$H_\ell(z \mid \text{tier} = k) = 1 - \bar{F}_{\tilde{S}}\left(\frac{z}{v} \mid \text{tier} = k\right) =$$

$$1 - \frac{1}{\mathbb{P}(\text{tier} = k)} \int_0^\infty \int_0^\pi 2\lambda_k r_0 \exp\left\{-\sum_{j\in\mathcal{K}} \lambda_j \left|\mathcal{A}_{kj}(r_0, \theta, z, 1, \beta_{kj})\right|\right\} d\theta dr_0. \tag{5.26}$$

To derive the chord length distribution in tier-k cells, according to (5.11), we need $\frac{d}{dz} H_\ell(z \mid \text{tier} = k)$ and $\mathbb{E}[L \mid \text{tier} = k]$. From (5.26), we have

$$\frac{d}{dz} H_\ell(z \mid \text{tier} = k) = \frac{1}{\mathbb{P}(\text{tier} = k)} \int_0^\infty \int_0^\pi 2\lambda_k r_0 \left(\sum_{j\in\mathcal{K}} \lambda_j \frac{d}{dz} \left|\mathcal{A}_{kj}(r_0, \theta, z, 1, \beta_{kj})\right|\right)$$

$$\times \exp\left\{-\sum_{j\in\mathcal{K}} \lambda_j \left|\mathcal{A}_{kj}(r_0, \theta, z, 1, \beta_{kj})\right|\right\} d\theta dr_0, \tag{5.27}$$

where $\frac{d}{dz}|\mathcal{A}_{kj}(r_0, \theta, z, 1, \beta_{kj})|$ is given in (5.28).

$\mathbb{E}[L \mid \text{tier} = k]$ is also provided in the following theorem.

Theorem 5.2. *The mean length of the chords lying in tier-k cells is as*

$$\mathbb{E}[L \mid \text{tier} = k] = \pi \frac{\left(\sum_{j\in\mathcal{K}} \lambda_j \beta_{jk}^2\right)^{1/2}}{\sum_{j\in\mathcal{K}} \lambda_j \mathcal{I}(\beta_{kj})},$$

where

$$\mathcal{I}(\beta) = \begin{cases} \frac{1}{\beta^2} \int_0^\pi \frac{\beta^2 + 1 - 2\cos^2\theta}{\sqrt{\beta^2 - \cos^2\theta}} d\theta, & \text{if } \beta \geq 1 \\ \frac{1}{\beta^2} \int_{\arccos(\beta)}^{\pi - \arccos(\beta)} \frac{\beta^2 + 1 - 2\cos^2\theta}{\sqrt{\beta^2 - \cos^2\theta}} d\theta, & \text{if } \beta < 1 \end{cases}. \tag{5.29}$$

Proof. See **Appendix C.3**. $\square$

$$\frac{\mathrm{d}}{\mathrm{d}z}|\mathcal{A}_{kj}(r_0, \theta, z, 1, \beta_{kj})| =$$

$$\begin{cases} \pi \frac{2(z-r_0\cos\theta)}{\beta_{kj}^2} + \frac{2}{\beta_{kj}^2} \int\limits_0^1 \sqrt{\beta_{kj}^2 g(z,u)^2 - (zu - r_0\cos\theta)^2} - \arccos\left(\frac{zu - r_0\cos\theta}{\beta_{kj} g(z,u)}\right)(zu - r_0\cos\theta)\mathrm{d}u \\ \quad + \frac{2z}{\beta_{kj}^2} \int\limits_0^1 \sqrt{\beta_{kj}^2 g(z,u)^2 - (zu - r_0\cos\theta)^2}\frac{u(zu - r_0\cos\theta)}{g(z,u)^2} - \arccos\left(\frac{zu - r_0\cos\theta}{\beta_{kj} g(z,u)}\right)u\mathrm{d}u, \\ \hfill \text{if } (\beta_{kj} > 1) \\[4pt] \pi\frac{2(z-r_0\cos\theta)}{\beta_{kj}^2} + \frac{\frac{\beta_{kj}^2-1}{\beta_{kj}^2}r_0^2}{\sqrt{4\beta_{kj}^2 r_0^2 - \left((\beta_{kj}^2-1)z + 2r_0\cos\theta\right)^2}} + \frac{\frac{\beta_{kj}^2+1-2\cos^2\theta}{\beta_{kj}^2}r_0^2 - \frac{\beta_{kj}^2-1}{\beta_{kj}^2}r_0 z\cos\theta}{\sqrt{4\beta_{kj}^2 g(z,1)^2 - \left((\beta_{kj}^2+1)z - 2r_0\cos\theta\right)^2}} \\ \quad - \arccos\left(\frac{(\beta_{kj}^2+1)z - 2r_0\cos\theta}{2\beta_{kj}g(z,1)}\right)\frac{2(z-r_0\cos\theta)}{\beta_{kj}^2} + \sqrt{\frac{-(\beta_{kj}^2-1)z + 2r_0(\beta_{kj}-\cos\theta)}{(\beta_{kj}^2-1)z + 2r_0(\beta_{kj}+\cos\theta)}}\frac{(\beta_{kj}^2-1)z + r_0(\beta_{kj}+\cos\theta)}{2\beta_{kj}^2} \\ \quad + \sqrt{\frac{(\beta_{kj}^2-1)z + 2r_0(\beta_{kj}+\cos\theta)}{-(\beta_{kj}^2-1)z + 2r_0(\beta_{kj}-\cos\theta)}}\frac{-(\beta_{kj}^2-1)z + r_0(\beta_{kj}-\cos\theta)}{2\beta_{kj}^2} \\ \hfill \text{if } (\beta_{kj} = 1) \text{ or } \left(\beta_{kj} < 1 \text{ and } 2r_0\frac{\cos\theta - \beta_{kj}}{1-\beta_{kj}^2} < z < 2r_0\frac{\cos\theta + \beta_{kj}}{1-\beta_{kj}^2}\right) \\[4pt] 0, \hfill \text{if } \left(\beta_{kj} < 1 \text{ and } z \le 2r_0\frac{\cos\theta - \beta_{kj}}{1-\beta_{kj}^2}\right) \\[4pt] \pi\frac{2(z-r_0\cos\theta)}{\beta_{kj}^2}, \hfill \text{if } \left(\beta_{kj} < 1 \text{ and } 2r_0\frac{\cos\theta + \beta_{kj}}{1-\beta_{kj}^2} \le z\right) \end{cases} \tag{5.28}$$

Using these results, the mean and the CCDF of the sojourn time are

$$\mathbb{E}[S \mid \text{tier} = k] = \frac{\pi}{v}\frac{\left(\sum_{j\in\mathcal{K}}\lambda_j\beta_{jk}^2\right)^{1/2}}{\sum_{j\in\mathcal{K}}\lambda_j\mathcal{I}(\beta_{kj})}, \tag{5.30}$$

$$\bar{F}_S(T \mid \text{tier} = k) = \mathbb{E}[L \mid \text{tier} = k]\frac{\mathrm{d}}{\mathrm{d}z}H_\ell(z \mid \text{tier} = k)\Big|_{z=vT}. \tag{5.31}$$

5.5.2 Handoff Rate

In [87], rates of different handoff types in multi-tier cellular networks are provided. For $k, j \in \mathcal{K}$, the type k-j handoff rate H_{kj}, defined as the mean number of handoffs made from a tier-k cell to a tier-j cell in unit time, is[2]

$$H_{kj} = \frac{v}{\pi}\frac{\lambda_k\lambda_j\mathcal{F}(\beta_{kj})}{\left(\sum_{i\in\mathcal{K}}\lambda_i\beta_{ik}^2\right)^{3/2}}, \tag{5.32}$$

where

$$\mathcal{F}(\beta) = \frac{1}{\beta^2}\int_0^\pi \sqrt{\beta^2 + 1 - 2\beta\cos\theta}\,\mathrm{d}\theta. \tag{5.33}$$

Therefore,

$$H_k = \frac{v}{\pi}\lambda_k\frac{\sum_{j\in\mathcal{K}}\lambda_j\mathcal{F}(\beta_{kj})}{\left(\sum_{i\in\mathcal{K}}\lambda_i\beta_{ik}^2\right)^{3/2}} \tag{5.34}$$

is the mean number of handoffs from (to) a tier-k cell to (from) any other cell in the network.

---

[2] H_{kj} in (5.32) is obtained by further simplifying the result in [87]. Specifically, we have used $\beta_{ij} = \frac{\beta_{ik}}{\beta_{jk}}$ besides $\mathcal{F}(\frac{1}{\beta}) = \beta^3\mathcal{F}(\beta)$.
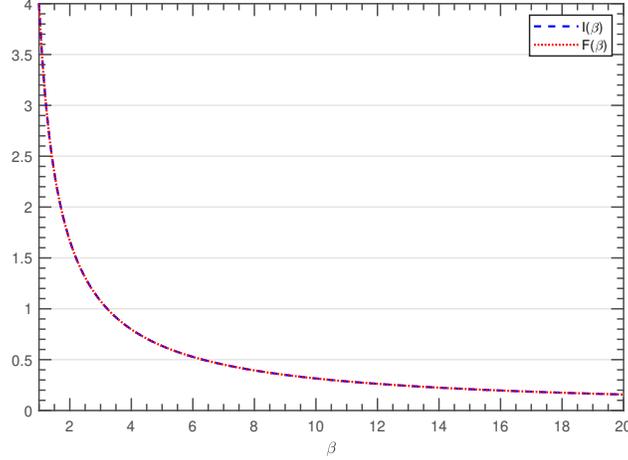
Figure 5.5: $\mathcal{I}(\beta)$ and $\mathcal{F}(\beta)$ for $\beta \geq 1$.

In Fig. 5.5, we can see $\mathcal{I}(\beta) = \mathcal{F}(\beta)$ for $\beta \geq 1$. Since $\mathcal{F}(\frac{1}{\beta}) = \beta^3 \mathcal{F}(\beta)$ and $\mathcal{I}(\frac{1}{\beta}) = \beta^3 \mathcal{I}(\beta)$, we can conclude $\mathcal{I}(\beta) = \mathcal{F}(\beta)$ for any $\beta > 0$. Using this result, the relation between mean sojourn time and handoff rate for tier-k is as follows:

$$\mathbb{E}[S \mid \text{tier} = k] = \frac{\mathbb{P}(\text{tier} = k)}{H_k}. \tag{5.35}$$

An important metric in mobility analysis is the fraction of time the mobile user stays in tier-k cells during a movement period since it considers both handoff rate and sojourn time.

Corollary 5.1. *The fraction of time the mobile user stays in tier-k, during the movement period, is $\mathbb{P}(\text{tier} = k)$.*

Proof. Let us denote the number of times that the mobile user enters a tier-k cell during the movement period by $\mathcal{N}_k$, $k \in \mathcal{K}$. Also, $t_i^{(k)}$ denotes the i-th dwell time in the tier-k cell, where $i = 1, ..., \mathcal{N}_k$ and $k \in \mathcal{K}$. The fraction of time that the mobile user stays in tier-k can be obtained by

$$\frac{\sum_{i=1}^{\mathcal{N}_k} t_i^{(k)}}{\sum_{j \in \mathcal{K}} \sum_{i=1}^{\mathcal{N}_j} t_i^{(j)}} = \frac{\sum_{i=1}^{\mathcal{N}_k} t_i^{(k)}}{\mathcal{N}_k} \times \frac{\mathcal{N}_k}{\sum_{j \in \mathcal{K}} \sum_{i=1}^{\mathcal{N}_j} t_i^{(j)}} = \mathbb{E}[S \mid \text{tier} = k] \times H_k \stackrel{(b)}{=} \mathbb{P}(\text{tier} = k),$$

where (b) follows from (5.35). □

5.5.3 Unconditional CCDF and Mean of Sojourn Time

So far we have focused on $\bar{F}_S(T \mid \text{tier} = k)$ and $\mathbb{E}[S \mid \text{tier} = k]$, i.e. the CCDF and average of sojourn time in a tier-k cell. In the following corollaries, we derive the unconditional CCDF and mean.

Corollary 5.2. *The CCDF of sojourn time, during the movement period, can be obtained by*

$$\bar{F}_S(T) = \sum_{k \in \mathcal{K}} \bar{F}_S(T \mid \text{tier} = k) \frac{H_k}{H},$$

where $H = \sum_{k \in \mathcal{K}} H_k$ is the mean number of handoffs in unit time.

Proof. Using the same notation as in the proof of **Corollary 5.1**, we can write

$$\bar{F}_S(T) = \mathbb{E}\left[\mathbf{1}(S > T)\right] = \frac{\sum_{k \in \mathcal{K}} \sum_{i=1}^{\mathcal{N}_k} \mathbf{1}(t_i^{(k)} > T)}{\sum_{j \in \mathcal{K}} \mathcal{N}_j}$$

$$= \sum_{k \in \mathcal{K}} \frac{\sum_{i=1}^{\mathcal{N}_k} \mathbf{1}(t_i^{(k)} > T)}{\mathcal{N}_k} \times \frac{\mathcal{N}_k}{\sum_{j \in \mathcal{K}} \sum_{i=1}^{\mathcal{N}_j} t_i^{(j)}} \times \frac{\sum_{j \in \mathcal{K}} \sum_{i=1}^{\mathcal{N}_j} t_i^{(j)}}{\sum_{j \in \mathcal{K}} \mathcal{N}_j}.$$

$\square$

Corollary 5.3. *The mean sojourn time during a movement period can be obtained by*

$$\mathbb{E}\left[S\right] = \frac{1}{H} = \frac{\pi}{v} \left(\sum_{k \in \mathcal{K}} \lambda_k \frac{\sum_{j \in \mathcal{K}} \lambda_j \mathcal{F}(\beta_{kj})}{\left(\sum_{i \in \mathcal{K}} \lambda_i \beta_{ik}^2\right)^{3/2}} \right)^{-1}.$$

Proof. The mean sojourn time can be obtained by following the same approach as in proof of **Corollary 5.2**.

$\square$

5.5.4 Effect of Network Parameters

In this subsection, we study the effect user speed, transmit power, bias factor, and BS intensity on the distribution and mean of the sojourn time. Since our proofs do not rely on the mobility model, the following results can be used for any mobility model.

Proposition 5.2. *The CCDF and the mean of the sojourn time decrease as the mobile user's speed increases.*

Proof. This can be understood from the definition of the sojourn time. (This can also be proven from the derived analytical results.)

$\square$

Proposition 5.3. *In multi-tier networks, the sojourn time of k-th tier increases as transmit power or bias factor of tier-k increases, while sojourn time in other tiers decreases. In single-tier networks, sojourn time is independent of transmit power and bias factor.*

Proof. Assume that a user at location $y \in \mathbb{R}^2$ is served by a tier-k BS at x, i.e, y is in the cell of x. Therefore,

from (5.1), we have

$$B_k P_k \|\mathrm{y} - x\|^{-\alpha} \geq B_k P_k \|\mathrm{y} - z\|^{-\alpha}, \qquad z \in \Phi_k, \tag{5.36}$$

$$B_k P_k \|\mathrm{y} - x\|^{-\alpha} \geq B_j P_j \|\mathrm{y} - z\|^{-\alpha}, \qquad z \in \Phi_j, \forall j \in \mathcal{K} \setminus \{k\}. \tag{5.37}$$

From these equations, when $B_k' P_k' \geq B_k P_k$, we obtain

$$B_k' P_k' \|\mathrm{y} - x\|^{-\alpha} \geq B_k' P_k' \|\mathrm{y} - z\|^{-\alpha}, \qquad z \in \Phi_k,$$

$$B_k' P_k' \|\mathrm{y} - x\|^{-\alpha} \geq B_j P_j \|\mathrm{y} - z\|^{-\alpha}, \qquad z \in \Phi_j, \forall j \in \mathcal{K} \setminus \{k\},$$

i.e. y is still in the cell of x after increasing $B_k P_k$. On the other hand, when the user at location y is served by a tier-j BS at x, $j \in \mathcal{K} \setminus \{k\}$, we have

$$B_k P_k \|\mathrm{y} - z\|^{-\alpha} \leq B_j P_j \|\mathrm{y} - x\|^{-\alpha},$$

where z denotes the location of the nearest tier-k BS to the user at y. For $B_k' P_k' \geq B_k P_k$, one of the two following cases can happen: a) $B_k' P_k' \|\mathrm{y} - z\|^{-\alpha} \leq B_j P_j \|\mathrm{y} - x\|^{-\alpha}$, i.e., the user is still served by the tier-j BS, or b) $B_j P_j \|\mathrm{y} - z\|^{-\alpha} \leq B_k' P_k' \|\mathrm{y} - x\|^{-\alpha}$, i.e., the nearest tier-k BS serves the user when $B_k P_k$ is increased to $B_k' P_k'$. Therefore, in multi-tier networks, the size of the k-th tier cells increases as transmit power or bias factor of tier-k increases while the size of other tiers' cells decreases. For single-tier networks, according to (5.36), cell sizes are independent of transmit power and bias factor. Finally, using these results and the fact that the sojourn time is directly proportional to the size of cells, we can obtain **Proposition 5.3**. $\qquad \square$

Proposition 5.4. *In multi-tier networks, when the BS intensity of tier-k increases, the sojourn time for other tiers decreases.*

Proof. According to the superposition property of PPP [41], increasing λ_k to λ_k' is similar to adding a new tier of BSs with intensity $\lambda_k' - \lambda_k$, transmission power P_k, and bias factor B_k. Therefore, the size of cells of other tiers decreases when the BS intensity of k-th tier increases. $\qquad \square$

5.5.5 Other Mobility Models

Due to irregular shapes of Poisson Voronoi cells and random direction changes between consecutive steps, analysis of sojourn time over multiple consecutive steps for mobility models such as random walk [86] or modified random waypoint [29] seems impossible. However, as the user velocity, flight length, or flight time

increases, the effect of direction change decreases. Thus, sojourn time approaches the results derived for straight line mobility model. In this regard, we have considered a simplified random walk mobility model where the mobile user moves with speed v for a duration of T_f before changing its direction. For small vT_f, irregular movements of the user near the cell boundaries can cause multiple handovers within a small period of time. On the other hand, when the user is far from the cell boundary it stays a longer time within the cell owing to the frequent random direction changes. For a two-tier network with $\lambda_1 = 0.002$, $\lambda_2 = 0.005$, and $\beta_{12} = \left(\frac{1}{2}\right)^{1/4}$, we have compared the mean sojourn time of the simplified random walk mobility model with our analytical results for different values of v and T_f in Table 5.2. It seems that the derived analytical results for the straight line mobility model can also be used for random mobility models when $\mathbb{E}[L] \leq vT_f$, where $\mathbb{E}[L]$ is the average chord length which is equal to $\sum_{k \in \mathcal{K}} \frac{\lambda_k}{\sum_{j \in \mathcal{K}} \lambda_j} \mathbb{E}[L \mid \text{tier} = k]$. For the given parameters, average chord length is about 8.

Table 5.2: Mean Sojourn Time

	$v = 0.5$	$v = 1$	$v = 4$
Analysis	18.74	9.37	2.34
$T_f = 1$	16.66	8.85	2.34
$T_f = 5$	18.49	9.24	2.35
$T_f = 10$	18.58	9.36	2.35

5.6 Numerical and Simulation Results

5.6.1 Distribution of Sojourn Time

For a two-tier cellular network, in Fig. 5.6(a), the distribution of the sojourn time $\tilde{S}$ in the cell where the connection is initiated is illustrated for tier-k, $k \in \{1, 2\}$. In Fig. 5.6(b), the distribution of (conditional and unconditional) S, for this network, is provided. As can be seen, the simulation results match the derived analytical results. According to Fig. 5.6, at high velocities, the sojourn time for tier-k stochastically dominates the sojourn time of tier-j when $B_k P_k > B_j P_j$, i.e. $\beta_{kj} > 1$.

In practice, when a mobile user crosses a cell boundary, it starts a Time to Trigger (TTT) timer. The mobile user does not make a handoff to the new BS, if it leaves the new BS's cell before the end of TTT timer [27]. The derived results for handoff rates, in the literature, usually assume TTT is 0, i.e. the handoff rates provided (in [29] and [87] for example) are actually the mean number of intersections between the user trajectory and cell boundaries per unit time. In practice, the handoff rate for tier-k is $H_k \mathbb{P}(S > \text{TTT} \mid \text{tier} = k) = H_k \bar{F}_S(\text{TTT} \mid \text{tier} = k)$, where H_k is given in (5.34). When the network parameters are as in Fig. 5.6, $H_1 = 0.13$ and $H_2 = 0.41$. For this network, with TTT $= 0.2$, the handoff rate for tier-one is 0.12 and for

(a) CCDF of $\tilde{S}$.   (b) CCDF of S.
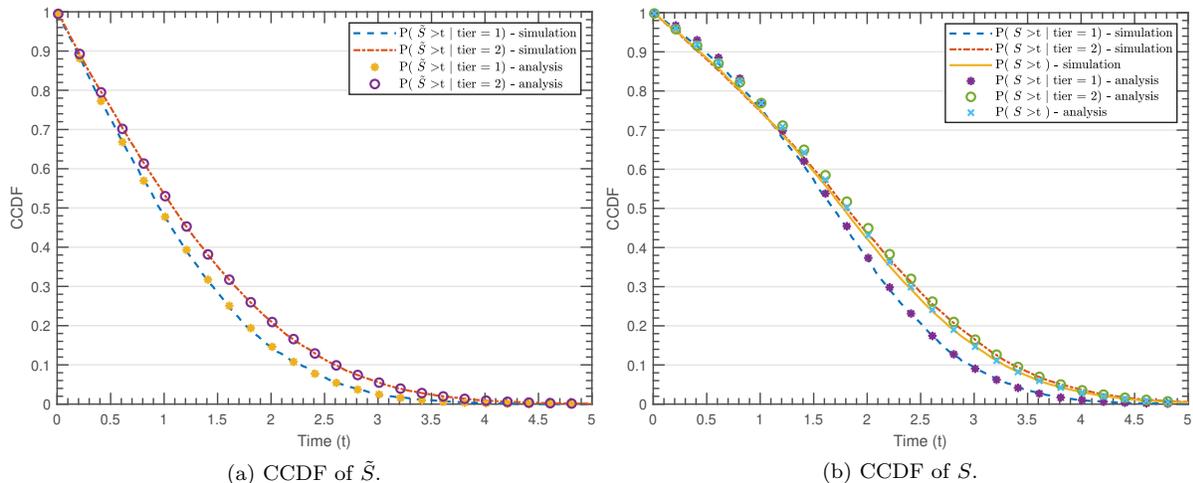
Figure 5.6: Distribution of $\tilde{S}$ and S in a two-tier cellular network (for $\lambda_1 = 0.002$, $\lambda_2 = 0.005$, $\beta_{12} = \left(\frac{1}{2}\right)^{1/4}$, and $v = 5$).

tier-two is 0.39. Although the difference between H_k and the handoff rate for these parameters is negligible, it is noticeable for high velocity scenarios. In this regard, for the same network parameters as in Fig. 5.6, effect of TTT on the handoff rates as a function of velocity is shown in Fig. 5.7. While H_k (mean number of intersections between the user's trajectory and tier-k cell boundaries per unit time) linearly increases with the user's velocity, handoff rate for tier-k ($H_k \mathbb{P}(S > \text{TTT} \mid \text{tier} = k)$) is close to zero at high velocities due to small sojourn time.

Moreover, using the distribution of sojourn time, we can study the ping-pong rate (unnecessary handoff rate). If, after a handoff, the time duration that the mobile user is inside the new cell be less than a threshold T_p, the handoff is considered unnecessary [27]. Therefore, for tier-k, the ping-pong rate can be obtained by

$$H_k \left(\mathbb{P}(S < T_p \mid \text{tier} = k) - \mathbb{P}(S < \text{TTT} \mid \text{tier} = k) \right) = H_k \left(\mathbb{P}(S > \text{TTT} \mid \text{tier} = k) - \mathbb{P}(S > T_p \mid \text{tier} = k) \right).$$

For TTT $= 0.2$ and $T_p = 0.5$, for tier-one, the ping-pong rate is less than 0.01 and for tier-two, the ping-pong rate is 0.03.

As discussed earlier, we obtain the distribution of $\tilde{S}$ in multi-tier networks from (5.2). For convex cells, (5.2) can be further simplified as (5.3) which is the complement of the handoff probability. Therefore, we can use (5.3) to derive the CCDF of the sojourn time in single-tier networks and also in multi-tier networks for the tier with the smallest BP (multiplication of bias factor and transmission power). However, for other tiers in multi-tier networks, (5.3) provides an upper bound for the CCDF of the sojourn time. This is also illustrated in Fig. 5.8 for a three-tier network. It is worth mentioning that the gap between the CCDF of
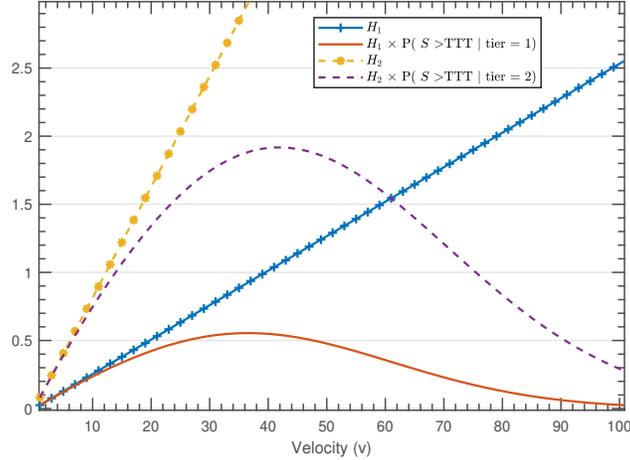
Figure 5.7: Effect of TTT on handoff rates for a two-tier cellular network with $\lambda_1 = 0.002$, $\lambda_2 = 0.005$, $\beta_{12} = \left(\frac{1}{2}\right)^{1/4}$, and TTT = 0.2.
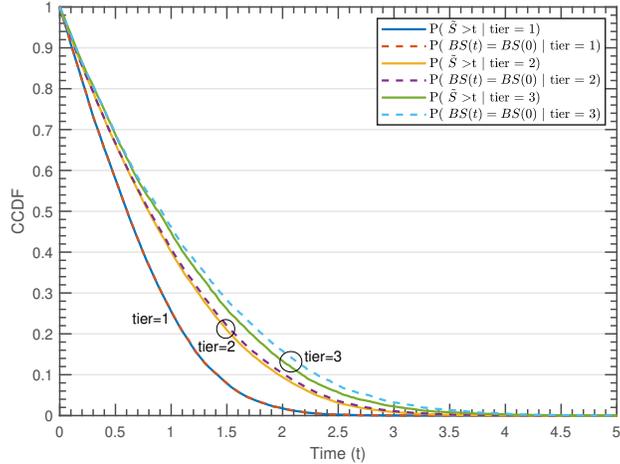
Figure 5.8: Sojourn time and complement of the handoff probability in a three-tier network. For tier-1, $\lambda_1 = 0.01$ and $B_1 P_1 = 10$, for tier-2, $\lambda_2 = 0.005$ and $B_2 P_2 = 50$, and for tier-3, $\lambda_3 = 0.001$ and $B_3 P_3 = 100$. $\alpha = 4$ and $v = 5$.

the sojourn time and its upper bound (obtained from (5.3)) increases as the intensity of tiers with lower BP increases.

5.6.2 Mean Sojourn Time

In Fig. 5.9, the mean sojourn time for a two-tier cellular network with $\lambda_1 = 0.002$ and $\lambda_2 = 0.005$ is illustrated as a function of velocity. We compare the results for the two-tier network with two single-tier scenarios where the mobile user is associated to only one of the tiers. When the number of tiers increases, through increased spectral reuse, users can transmit with higher data rates. However, there is more undesired overhead transmission due to the higher handoff rate (lower mean sojourn time). The sojourn

79
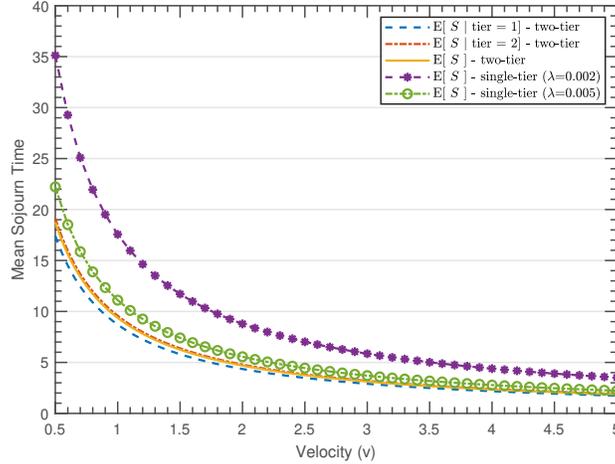
Figure 5.9: Mean sojourn time with respect to velocity. We have compared the results for a two-tier cellular network with $\lambda_1 = 0.002$, $\lambda_2 = 0.005$, and $\beta_{12} = \left(\frac{1}{2}\right)^{1/4}$ with two single-tier networks.

time distribution is helpful in mobility management where the mobile user can skip unnecessary handoffs with a negligible spectral efficiency loss [91].

In Fig. 5.10(a), the effect of transmit power (or bias factor) on the mean sojourn time in a two-tier cellular network is illustrated. As discussed in **Proposition 5.3**, the mean sojourn time of tier-two increases as transmit power (or bias factor) of tier-two increases, while mean sojourn time of other tier decreases. As can be seen, the (unconditional) mean sojourn time in the network does not change with increasing transmit power or bias factor of tier-two. In Fig. 5.10(b), the effect of BS intensity on the mean sojourn time is shown. As can be seen, the mean sojourn time for all tiers decreases with increasing the BS intensity of tier-two. This is also mentioned in **Proposition 5.4**.

5.7 Conclusion

We have derived the distribution and mean of the sojourn time of multi-tier cellular networks. The existing works assume that a mobile user is always associated to only one of the tiers, or focus on the sojourn time in small cells (for two-tier scenario). Since in both the cases the cells are convex, the sojourn time distribution (or mean) can be easily obtained similar to single-tier scenarios by using the chord length distribution in Poisson Voronoi tessellation. However, in multi-tier networks with maximum biased averaged received power association we need the chord length distribution in weighted Poisson Voronoi tessellation, which is not available in the literature. In this chapter, we have derived the area of contact in weighted Poisson Voronoi tessellation from which we have obtained the linear contact distribution function and chord length distribution.

(a) Effect of transmission power (or bias factor).　　　　(b) Effect of BS intensity.
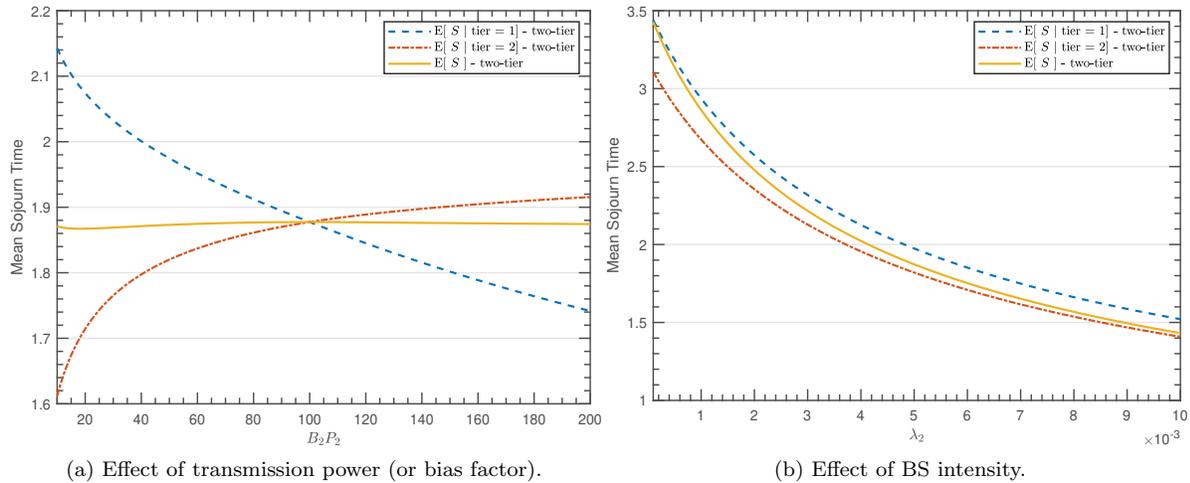
Figure 5.10: Effect of network parameters on the mean sojourn time in a two-tier cellular network with $\lambda_1 = 0.002$, $B_1 P_1 = 100$, $\alpha = 4$, and $v = 5$. a) Effect of increasing transmission power (or bias factor) when $\lambda_2 = 0.005$. b) Effect of increasing BS intensity when $B_2 P_2 = 50$.

We have studied the relation between mean sojourn time and other mobility-related performance metrics. Specifically, We have shown that mean sojourn time is inversely proportional to the handoff rate. Also, the complementary cumulative distribution function of sojourn time is upper bounded by complement of the handoff probability. In addition, we have studied the impact of user velocity and network parameters on the distribution and mean of the sojourn time. The sojourn time distribution can be used to derive the ping-pong rate which is important in mobility management where the mobile user can skip unnecessary handoffs with a negligible spectral efficiency loss. Moreover, it can be used for studying channel occupancy time which can be exploited for improving resource allocation.

Chapter 6

Federated Learning in Unreliable and Resource-Constrained Cellular Wireless Networks

With growth in the number of smart devices and advancements in their hardware, in recent years, data-driven machine learning techniques have drawn significant attention. However, due to privacy and communication issues, it is not possible to collect this data at a centralized location. Federated learning is a machine learning setting where the centralized location trains a learning model over remote devices. Federated learning algorithms can not be employed in the real world scenarios unless they consider unreliable and resource-constrained nature of the wireless medium. In this chapter, we propose a federated learning algorithm that is suitable for wireless networks. We prove its convergence, and provide a sub-optimal scheduling policy that improves the convergence rate. We also study the effect of local computation steps and communication steps on the convergence of our algorithm. We prove, in practice, federated learning algorithms may solve a different problem than the one that they have been employed for if they neglect the unreliability of wireless channels. Finally, through numerous experiments on real and synthetic datasets, we demonstrate the convergence of our proposed algorithm.

6.1 Introduction

6.1.1 Motivation

With the rapid growth in Internet-of-Things (IoT) applications and increase in the computational and storage power of smart devices, modern distributed networks generate a huge amount of data everyday [32]. Owing to this reason, data-driven machine learning techniques have gained significant attention in recent years. Currently, most of the existing machine learning techniques are centralized, i.e., they assume all data is available at a centralized location, where a central processor trains a powerful learning method on the data [30]. However, transferring data from user devices to the centralized location violates users' privacy [33]. To cope with this issue, federated learning has been introduced where a learning model is trained over remote devices under the control of the centralized location, called server [32, 33, 34]. Specifically, in federated learning (FL), the server broadcasts the global model parameters to the remote devices. Each remote device uses its local dataset to update the global model, and then transmits the updated local model to the server. After aggregating the local models, the server updates the global model and repeats the whole procedure. As an example, consider the task of next-word prediction on mobile phones, where a language model predicts the most probable next word or phrase based on a small amount of user-generated preceding text. To maintain the users privacy, instead of transmitting the raw text data to the server and training a predictor at the server, we use federated learning [100, 32].

Although the above definition of federated learning seems similar to parallel optimization and distributed machine learning in datacenters, due to the following challenges, it needs to be treated separately: i) In federated learning, connections between the remote devices and the server[1] are unreliable and slow. ii) Different devices in the network have different systems characteristics (systems heterogeneity). iii) Training data is not independently and identically distributed (statistical heterogeneity) [37, 38]. Since data is non-i.i.d. (not independently and identically distributed) across devices, all devices must participate in the learning process. However, due to the systems heterogeneity, which includes variable computation and communication capabilities at different devices, and limited number of available resource blocks (such as bandwidth) for communication, full device participation at each round of communication is not possible. Moreover, in reality, transmission success probability is different for different devices, even when they all have the same hardware. Specifically, transmission success probability for devices that are located closer to the server is higher than far devices. Thus, without considering this issue at the time of updating the global model, the updated global model will be biased towards cell center devices' local models.

---

[1]In wireless networks, base stations (BSs) act as the servers.

6.1.2 Related Works

Training federated learning models in a wireless networking environment requires devices and the server exchange information via wireless transmissions. The existing works (e.g. in [33, 101], and references therein) aim to improve implementation of federated learning by optimizing resource allocation and/or reducing the communication requirements since communication is a key bottleneck [32]. To reduce the communication rounds of FL, three main approaches have been employed: quantization, sparsification, and local updates [102]. In this chapter, we study local updates where, between any two aggregation steps in consecutive rounds, each device performs multiple local update steps. In this regard, [30, 31, 103] studied the convergence of communication-efficient FL with local updates for convex and non-convex problems. However, it was assumed that all of the devices participate in the aggregation step, which is obviously not possible when the number of available resource blocks is limited. To tackle this problem, in [35, 36, 37, 38, 39, 40] at the beginning of each round, the server samples a subset of devices and allocates the available resource blocks to these devices. After performing local update steps, the BS aggregates the local models of the chosen (scheduled) devices and updates the global model. The works in [35, 36, 37, 38, 39, 40] assumed that the BS successfully receives the local models of all the scheduled devices, and they designed the global model update step only based on the scheduling policy. However, in reality, not only is the success probability less than one, but also it is different for different devices. Clearly, to update the global model we need to include both scheduling policy and success probability.

To better understand, consider a scenario with two devices and two resource blocks, where both devices are scheduled for sharing their updated local models with the server. Assume that the success probability is 1 for *device one* and is 0.1 for *device two*. In this scenario, the server receives the local model of *device two* once in every ten rounds on average, while the local model of *device one* is always successfully received. Obviously, without considering this aspect of wireless communications, the global model is biased towards *device one* at the end of the learning process.

In the above context, we propose an FL algorithm that is suitable for unreliable and resource-constrained cellular wireless networks. Specifically, at the end of each FL round, our algorithm updates the global model by allocating different weights to different devices, where weight for each device depends on the scheduling policy and its transmission success probability. Stochastic geometry is used in the literature to derive the transmission success probabilities in large-scale cellular networks ([21, 43, 46, 47] and references therein). Therefore, in designing FL algorithms for cellular wireless networks, we need to borrow tools from stochastic geometry.

To the best of our knowledge, only the work in [104] considered the effect of success probability on the

convergence of FL. Similar to [105], [104] solved the FL problem using primal-dual optimization method. However, when strong duality is not guaranteed, this method may not be useful [36]. Also, in the analysis, the FL global objective is assumed to be a function of linear combination of model parameters and the input features. Thus, convergence analysis for this method cannot be extended to other machine learning techniques.

6.1.3 Contributions

The most common federated learning algorithm in the literature is FedAvg (federated averaging) [35], which lacks convergence analysis. Several works in the literature made steps towards analyzing convergence of FedAvg by modifying the algorithm. For example, [37] proposed a different averaging scheme for FedAvg, and derived the convergence rate of the modified algorithm. [36] proposed FedProx, which is a generalization of FedAvg obtained by adding a proximal term to the local objectives. Recently, [38] has proposed SCAFFOLD which improves FedAvg by adding a correction term, called client drift, to local updates. However, as we discussed in the previous subsection, these works and all other related works do not consider the effect of transmission success probability.

In the above context, the major contributions of this chapter can be summarized as follows:

- We propose an FL algorithm that is suitable for unreliable and resource-constrained wireless systems. In particular, for an FL system in a cellular wireless network, the BS selects a subset of devices at each round and updates the global model based on their updated local models by allocating different weights to them. Weight for each device depends on the scheduling policy and its transmission success probability. We use stochastic geometry tools to approximately calculate the success probability for each device.

- Our proposed FL algorithm solves the FL problem in the primal domain. We prove that, for strongly convex and smooth problems, the algorithm converges on non-i.i.d data with rate $\mathcal{O}(\frac{1}{T})$.

- We study two difference scheduling (sampling) policies. We also provide a sub-optimal scheduling policy based on the derived convergence rate.

- We study the effect of number of computation steps and communication steps on the convergence rate.

- We show that the existing works, which do not include the transmission success probability in the global model update step (e.g. those in [35, 36, 37, 38]), will not be suitable for a wireless communication environment, since they may converge to the solution of a different FL problem when the success probabilities are different for different devices.

- We verify the convergence of our proposed FL algorithm by experimenting over real and synthetic datasets. We also compare our results with centralized full batch gradient descent which can be considered as a benchmark for our algorithm.

The organization of the rest of the chapter is as follows. In Section 6.2, we introduce the system model and propose our FL algorithm. Then, in Section 6.3, we provide the convergence analysis of our FL algorithm. Further analysis of the proposed FL algorithm and comparison with related algorithms are provided in Section 6.4. In Section 6.5, we present the simulation results. Finally, Section 6.6 concludes the chapter. A summary of the major notations used in this chapter is given in Table 6.1.

Table 6.1: Summary of Major Notations

Notation	Description
λ	Base station (BS) intensity (average number of BSs in a unit area)
N	Number of devices in each cell
M	Number of available resource blocks at each BS
F, F_k	Global loss function, local loss function at device k
w, w^k	Global model parameters, local model parameters at device k
p_k	Weight of k-th nearest device
q_k	Average number of allocated resource blocks to device k at a sampling step
U_k	Success probability of device k
K	Total number of FL rounds (iterations)
E	Number of local SGD steps during each round of FL
ℓ	Number of transmission attempts at each aggregation step

6.2 System Model

6.2.1 Network Model

Consider a single-tier cellular network. The Locations of the base stations (BSs) follow a homogeneous Poisson point process (PPP) Φ of intensity λ, and each BS serves the user devices that are located in its Voronoi cell. Each device is associated to its nearest BS. In each cell, N devices are uniformly distributed; we use subscript k to denote the k-th nearest device to the BS. Thus, r_k denotes the distance between the serving BS and its k-th nearest device. Each BS allocates M resource blocks for the learning process, where $M \leq N$.
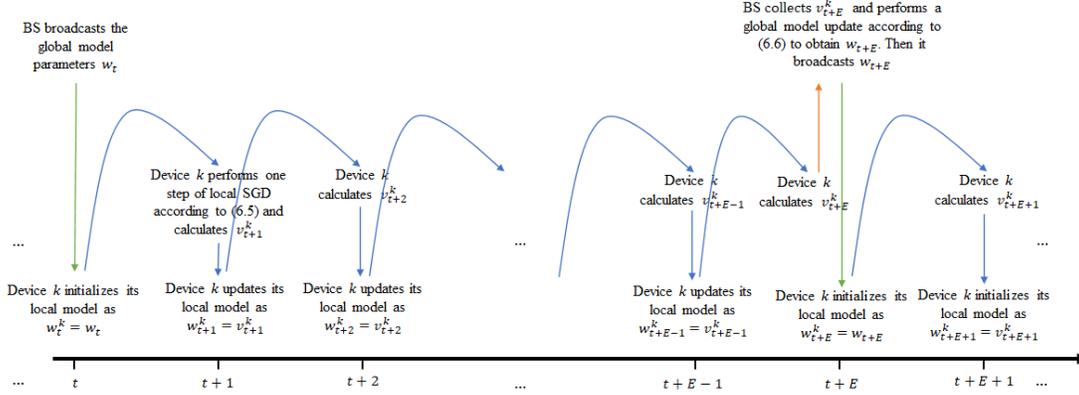
Figure 6.1: One iteration (round) of the distributed learning algorithm starting at time t. Green arrows illustrate the broadcast step. Blue arrows correspond to the local SGD step, and red arrow represents the aggregation step.

6.2.2 Federated Learning

In order to learn a statistical model from the distributed data across user devices, BS tries to solve the following distributed optimization problem:

$$\min_{w} \quad F(w) = \sum_{k=1}^{N} p_k F_k(w), \tag{6.1}$$

where w is the learning model parameters. p_k is the weight of the k-th nearest device such that $p_k \geq 0$ and $\sum_{k=1}^{N} p_k = 1$. $F_k(w)$ also denotes the local loss function at device k; it is defined as

$$F_k(w) = \frac{1}{n_k} \sum_{x \in \mathcal{D}_k} \mathcal{L}(w, x), \tag{6.2}$$

where $\mathcal{D}_k$ is the local dataset at device k, and is non-i.i.d. across different devices. $n_k = |\mathcal{D}_k|$ denotes the number of samples in $\mathcal{D}_k$. Thus, in (6.1), we can set $p_k = \frac{n_k}{n}$, where $n = \sum_{k=1}^{N} n_k$. $\mathcal{L}(w, x)$ also represents the loss function for data sample x.

Since the information is distributed across multiple devices, BS can not directly solve (6.1). Therefore, BS and devices collaboratively learn the optimal model parameters w^* [2] by following an iterative algorithm. Specifically, after initializing the model parameters at the BS at time 0, each iteration (round) of the algorithm comprises: 1) *Sampling and broadcast*, 2) *Local stochastic gradient descent (SGD)*, and 3) *Aggregation and averaging*. In the following we discuss each of these steps in detail for an iteration that starts at time t. This iteration is also shown in Fig. 6.1.

Sampling and Broadcast: Due to the limited number of available resource blocks, the BS first selects

---

[2] $w^* = \arg\min F(w)$

87

a group of devices at time t and then broadcasts its model parameters. In this chapter, we consider two different sampling schemes.

In *Scheme I*, BS uniformly selects M devices out of N devices without replacement, i.e., each device uses at most one resource block. Lets denote the set of the selected (scheduled) devices at time t by $\mathcal{S}_t$. For *Scheme I*, device k uses q_k resource blocks on average at sampling time t, where

$$q_k = \mathbb{E}\left[\sum_{m=1}^{M} \mathbf{1}\left(k \in \mathcal{S}_t(m)\right)\right] = \frac{M}{N}. \tag{6.3}$$

$\mathcal{S}_t(m)$ in the above equation denotes the index of the scheduled device at time t for resource block m, and $\mathbf{1}(.)$ is the indicator function.

In *Scheme II*, at time t for resource block m, BS samples a device with replacement from N devices with probabilities $\{\hat{q}_k\}$. Thus, sampling is independently and identically distributed over m, and some devices may use more that one resource block. For *Scheme II*, we have

$$q_k = \mathbb{E}\left[\sum_{m=1}^{M} \mathbf{1}\left(k \in \mathcal{S}_t(m)\right)\right] = M\hat{q}_k. \tag{6.4}$$

Local SGD: After receiving the global model parameters from the BS at time t, scheduled device k initializes its local model as $w_t^k = w_t$, where w_t is the global model parameters sent by the BS, and w_t^k is the local model parameters at device k. Then device k performs E steps of SGD and updates its local model at each step as follows

$$v_{t+i+1}^k = w_{t+i}^k - \eta_{t+i}\nabla F_k(w_{t+i}^k; \xi_{t+i}^k), \qquad i = 0, ..., E-1, \tag{6.5}$$

where η_{t+i} is the learning rate, and ξ_{t+i}^k is the sample uniformly chosen from the local dataset at device k[3]. Moreover,

$$w_{t+i}^k = \begin{cases} w_t & i = 0 \\ v_{t+i}^k & i = 1, 2, ..., E-1, \end{cases}.$$

Aggregation and Averaging: At the end of E steps of stochastic gradient descent, scheduled devices transmit $v_{t+E}^k - w_t$ to the BS using the allocated resource block(s). After collecting the successfully received local models, the BS updates the global model as

$$w_{t+E} = w_t + \sum_{k=1}^{N}\sum_{m=1}^{M} \frac{p_k}{q_k U_k}\mathbf{1}\left(k \in \mathcal{S}_t(m), \mathrm{SINR}_{k,m} > \theta\right)(v_{t+E}^k - w_t), \tag{6.6}$$

---

[3]In later sections, we discuss generalizing our results to batch size of greater than one.

where $\mathrm{SINR}_{k,m}$ denotes the recieved signal-to-interference-plus-noise ratio of device k over resource block m. θ denotes the SINR threshold, and U_k is defined as

$$U_k = \mathbb{E}\left[\mathbf{1}\left(\mathrm{SINR}_{k,m} > \theta\right) \mid k \in \mathcal{S}_t(m)\right],$$

i.e., U_k denotes the success probability of device k given that it is scheduled to use resource block m.

According to (6.6), the BS updates the global model by allocating different weights to different devices. The weight for device k is a function of scheduling policy (q_k) and transmission success probability (U_k) to compensate for systems heterogeneity and different communication capabilities. However, the averaging step in previous works, such as [36, 37, 40], is designed only based on the scheduling policy while assuming that success probabilities are all equal to one. As a result, the algorithms proposed in [36, 37, 40] (and other similar works) do not solve (6.1) in real world deployments, and this will be explained in Section 6.4.3

Remark: From (6.5), we have

$$v_{t+E}^k - w_t = -\sum_{i=t}^{t+E-1} \eta_i \nabla F_k(w_i^k; \xi_i^k).$$

Thus, instead of transmitting $v_{t+E}^k - w_t$, device k can directly send the gradients (and also the learning rates). However, direct transmission of the gradients requires sending E times more parameters than $v_{t+E}^k - w_t$.

Based on the above discussion, we have summarized our proposed FL algorithm in **Algorithm 1**.

Algorithm 1: Proposed FL Algorithm

Input: K, E
Initialization: Global model parameters at the BS w_0
for $i = 0, 1, ..., K-1$ **do**
> **Step 1 (Sampling and Broadcast):** BS samples a subset of devices and broadcasts the global model parameters w_{iE}.
> **Step 2 (Local SGD):** Each sampled device initializes its local model parameters with w_{iE}, and performs E steps of local SGD following (6.5) with $t = iE$.
> **Step 3 (Aggregation and Averaging):** Scheduled devices transmit their updated local models to the BS. Then, BS updates the global model parameters according to (6.6) and obtains $w_{(i+1)E}$.

return w_{KE}

Moreover, we can summarize the update rule of the proposed algorithm as

$$v_{t+1}^k = w_t^k - \eta_t \nabla F_k(w_t^k; \xi_t^k), \qquad \forall k \in \{1, ..., N\}, \tag{6.7}$$

where

$$w_t^k = \begin{cases} w_0 & t = 0 \\ v_t^k & t \notin \mathcal{I}_E \\ w_t = w_{t-E} + \sum_{k=1}^{N} \sum_{m=1}^{M} \frac{p_k}{q_k U_k} \mathbf{1}\left(k \in \mathcal{S}_{t-E}(m), \mathrm{SINR}_{k,m} > \theta\right)(v_t^k - w_{t-E}), & t \in \mathcal{I}_E \end{cases} \quad (6.8)$$

$\mathcal{I}_E = \{nE \mid n = 1, 2, ...\}$ is the set of time indexes of the global synchronization steps after $t = 0$ [37].

6.2.3 Interference, SINR, and Success Probability

In this subsection, we study the success probability of device k at its associated BS, given that it is scheduled to transmit over resource block m. Due to the stationarity of the homogeneous PPP [41], we can consider the location of this BS as the origin of our coordination system.

Before studying the uplink success probability, it is worth mentioning that, in this chapter, we have assumed downlink communication is always successful, i.e., all the scheduled devices successfully receive the global model parameters. This is a valid assumption since the BS can transmit with more power compared to the devices; moreover, the BS can allocate more than one resource block for broadcasting the global model parameters. However, due to the power constraint at user devices and limited number of allocated resource blocks to each scheduled device, we must consider that uplink transmissions may not be successful at an aggregation step.

To increase the success probability during the aggregation steps, we assume all scheduled devices transmit their local model parameters for ℓ times, and BS employs the selection combining, i.e., it uses the highest received SINR to recover the local model parameters transmitted over resource block m. Therefore,

$$U_k = \mathbb{E}\left[\mathbf{1}\left(\mathrm{SINR}_{k,m} > \theta\right) \mid k \in \mathcal{S}_t(m)\right]$$

$$= \mathbb{E}\left[\mathbf{1}\left(\max\left\{\mathrm{SINR}_{k,m}(1), \mathrm{SINR}_{k,m}(2), ..., \mathrm{SINR}_{k,m}(\ell)\right\} > \theta\right) \mid k \in \mathcal{S}_t(m)\right],$$

where $\mathrm{SINR}_{k,m}(i)$ denotes the received SINR from device k over resource block m at i-th transmission attempt of an aggregation step. Since the success event is identically distributed across different iterations (rounds) of the algorithm for device k, $\mathrm{SINR}_{k,m}(i)$ is not indexed by the iteration number. Moreover, at the aggregation step of each iteration, interference and fading are identically distributed across different resource blocks; thus, we can omit subscript m from $\mathrm{SINR}_{k,m}(i)$ since its statistics does not depend on m. With this in mind, we can write

$$\mathrm{SINR}_k(i) = \frac{h_k(i)r_k^{-\alpha}}{I(i) + \sigma^2} = \frac{h_k(i)r_k^{-\alpha}}{\sum_{x \in \Phi_I} h_x(i)\|x\|^{-\alpha} + \sigma^2}, \quad (6.9)$$

where σ^2 is the normalized noise power (noise power to the device transmit power). $I(i) = \sum_{x \in \Phi_I} h_x(i) \|x\|^{-\alpha}$ is the interference at i-th transmission attempt with Φ_I denoting the set of interferers. $h_x(i)$ denotes the small scale fading between a device at x and the BS at the origin. We consider Rayleigh fading, i.e., $h_x \sim \exp(1)$, and it is i.i.d. across i (different transmission attempts) and x (different locations). α also denotes the path-loss exponent. Since the set of interferers remains the same during different transmission attempts of one aggregation step, success events are temporally correlated over one resource block during an aggregation step.

It is also worth mentioning that our algorithm can also be employed with the results in [106] for maximal ratio combining. Moreover, we can easily extend our framework to more complex scenarios including uplink transmission with power control as in [107], multiple input multiple output antenna systems as in [108, 109], millimeter wave as in [110].

Lemma 6.1. *For the described network, with ℓ transmission attempts at each aggregation step, the success probability for scheduled device k is*

$$U_k \approx \sum_{i=1}^{\ell} \binom{\ell}{i} (-1)^{i+1} \exp\left\{ -i\theta\sigma^2 r_k^\alpha - 2\pi\lambda \int_0^\infty \left(1 - \frac{1}{(1 + \theta r_k^\alpha x^{-\alpha})^i} \right) (1 - e^{-12/5\lambda\pi x^2}) x \mathrm{d}x \right\}.$$

Proof. See **Appendix D.1**. □

6.3 Convergence Analysis of the Proposed Federated Learning Algorithm

6.3.1 Convergence Rate

In this section, we prove the convergence of the proposed algorithm. In this regard, we first introduce additional notations and assumptions that are required to derive the convergence rate.

Following the same notation as in [37], for time t, we define

- $\bar{v}_t = \sum_{k=1}^{N} p_k v_t^k$,

- $\bar{w}_t = \sum_{k=1}^{N} p_k w_t^k$,

- $g_t = \sum_{k=1}^{N} p_k \nabla F_k(w_t^k; \xi_t^k)$,

- $\bar{g}_t = \sum_{k=1}^{N} p_k \nabla F_k(w_t^k)$, where $\nabla F_k(w_t^k) = \mathbb{E}_\xi \left[\nabla F_k(w_t^k; \xi) \right]$.

Thus, at $t \in \mathcal{I}_E$, $w_t^k = \bar{w}_t = w_t$.

We also make the following assumptions [37, 111, 112].

Assumption 6.1. $F_1, ..., F_N$ are all μ-strongly convex, i.e., for $k \in \{1, ..., N\}$,

$$F_k(w_2) \geq F_k(w_1) + \nabla F_k(w_1)^T (w_2 - w_1) + \frac{\mu}{2} \|w_2 - w_1\|^2, \qquad \forall w_1, w_2.$$

Consequently, F is μ-strongly convex.

Assumption 6.2. $F_1, ..., F_N$ are all L-smooth, i.e., for $k \in \{1, ..., N\}$,

$$F_k(w_2) \leq F_k(w_1) + \nabla F_k(w_1)^T (w_2 - w_1) + \frac{L}{2} \|w_2 - w_1\|^2, \qquad \forall w_1, w_2.$$

Consequently, F is L-smooth.

Assumption 6.3. *Variance of stochastic gradient at device k, $k \in \{1, ..., N\}$, is upper bounded by σ_k^2, i.e.,*

$$\mathbb{E}_\xi \left[\|\nabla F_k(w; \xi) - \nabla F_k(w)\|^2 \right] \leq \sigma_k^2, \qquad \forall w.$$

Assumption 6.4. *Second moment of the norm of the stochastic gradient is bounded at all devices. For all $k \in \{1, ..., N\}$,*

$$\mathbb{E}_\xi \left[\|\nabla F_k(w; \xi)\|^2 \right] \leq G^2, \qquad \forall w.$$

We also define $\Gamma = F^* - \sum_{k=1}^N p_k F_k^*$, where F^* is the minimum global loss (objective function in (6.1)) and F_k^* is the minimum local loss at device k (objective function in (6.2)). $\Gamma \geq 0$ (proof is given bellow), and increases as the heterogeneity (degree of non-i.i.d.) of the data distribution increases [37].

Proof of $\Gamma \geq 0$. As mentioned previously, w^* minimizes the global loss, i.e. $F^* = F(w^*)$. Similarly, we define $w_k^* = \arg\min F_k(w)$. From the definition of Γ, we have

$$\Gamma = F(w^*) - \sum_{k=1}^N p_k F_k(w_k^*) \overset{(a)}{=} \sum_{k=1}^N p_k F_k(w^*) - \sum_{k=1}^N p_k F_k(w_k^*) = \sum_{k=1}^N p_k \left(F_k(w^*) - F_k(w_k^*) \right) \overset{(b)}{\geq} 0,$$

where (a) follows from the definition of the global loss function, and (b) is obtained since w_k^* minimizes F_k. When data is i.i.d. across devices, $F_1 = F_2 = \cdots = F_N = F$. Thus, $\Gamma = 0$ for i.i.d. data distribution. $\qquad \square$

Theorem 6.1. *When $\eta_t = \frac{2}{\mu(\gamma+t)}$ with $\gamma = \max\left\{8\frac{L}{\mu}, E\right\}$, after the averaging step at time T, we have*

$$\mathbb{E}[F(w_T) - F^*] \leq \frac{L/\mu}{\gamma + T} \left[\frac{2}{\mu} \left(\sum_{k=1}^N p_k^2 \sigma_k^2 + 6L\Gamma + 8(E-1)^2 G^2 + 4E^2 G^2 B \right) + \frac{\mu\gamma}{2} \|w_0 - w^*\|^2 \right],$$

where $B = \sum_{k=1}^{N} p_k \left(\frac{1}{q_k U_k} - 1 \right)$ *for sampling Scheme I, and* $B = \sum_{k=1}^{N} p_k \left(\frac{1}{q_k U_k} - \frac{1}{M} \right)$ *for sampling Scheme II.* w_0 *denotes the initialized global model parameters.*

Proof. The result is obtained from

$$\mathbb{E} \left[\| \bar{w}_{t+1} - w^* \|^2 \right] \leq (1 - \mu \eta_t) \mathbb{E} \left[\| \bar{w}_t - w^* \|^2 \right] + \eta_t^2 \left(\underbrace{\sum_{k=1}^{N} p_k^2 \sigma_k^2}_{\text{term I}} + \underbrace{6L\Gamma}_{\text{term II}} + \underbrace{8(E-1)^2 G^2}_{\text{term III}} + \underbrace{4E^2 G^2 B}_{\text{term IV}} \right). \quad (6.10)$$

For details, see **Appendix D.2**. □

[37] established a convergence rate of $\mathcal{O} \left(\frac{1}{T} \right)$ based on **Assumptions 1 to 4**. According to **Theorem 1**, our proposed federated learning algorithm also converges with rate $\mathcal{O} \left(\frac{1}{T} \right)$ when **Assumptions 1 to 4** hold. Thus, unsuccessful transmissions do not affect the convergence significantly after proper adjustment of the averaging step.

In (6.10), **term I** is due to the fact that each device uses a mini-batch instead of the full batch to perform a local update, **term II** exists because data is non-i.i.d. across devices, **term III** stems from performing multiple $(E > 1)$ local SGD steps between two aggregation steps which allows local models to move without the server's control in the direction of the local optimum instead of the global optimum, and **term IV** comes from the transmission unreliability and resource scarcity in cellular wireless networks. When **term I**, **term II**, **term III**, and **term IV** are zero, according to (6.10), the global model converges exponentially fast towards w^*. This is also the case in centralized full batch gradient descent. Thus, in Section 6.5, we use centralized full batch gradient descent as a benchmark.

So far we have assumed that each device uses only one data sample at each local update step. In the following, we discuss using mini-batches with more than one data sample at the local update steps.

6.3.2 Mini-Batch Gradient Descent

When we use mini-batches with size b for local update steps, (6.7) changes to

$$v_{t+1}^k = w_t^k - \eta_t \nabla F_k(w_t^k; \{\xi_t^k\}), \qquad \forall k \in \{1, ..., N\}, \quad (6.11)$$

where

$$\nabla F_k(w_t^k; \{\xi_t^k\}) = \frac{1}{b} \sum_{\xi \in \mathcal{B}_t^k} \nabla F_k(w_t^k; \xi)$$

93

is the estimated gradient at device k at time t using samples in the mini-batch $\mathcal{B}_t^k = \{\xi_t^k\}$. For device k and $\forall w$, we have

$$
\mathbb{E}_{\{\xi\}}\left[\left\|\nabla F_k(w;\{\xi\}) - \nabla F_k(w)\right\|^2\right] = \mathbb{E}_{\{\xi\}}\left[\left\|\sum_{\xi\in\mathcal{B}^k}\frac{1}{b}\left(\nabla F_k(w;\xi) - \nabla F_k(w)\right)\right\|^2\right]
$$
$$
\leq \mathbb{E}_{\{\xi\}}\left[\sum_{\xi\in\mathcal{B}^k}\frac{1}{b}\left\|\left(\nabla F_k(w;\xi) - \nabla F_k(w)\right)\right\|^2\right] \overset{(a)}{\leq} \sigma_k^2,
$$

where (a) is obtained from **Assumption 6.3**. Similarly, from **Assumption 6.4**, we have

$$
\mathbb{E}_{\{\xi\}}\left[\left\|\nabla F_k(w;\{\xi\})\right\|^2\right] \leq G^2, \qquad \forall w.
$$

Therefore, **Theorem 6.1** holds for any batch size.

6.4 Further Analysis and Comparison

In this section, we provide a further discussion on the proposed federated learning algorithm using **Theorem 6.1**. Specifically, we first study the impact of number of iterations (rounds), number of local update steps at each round, and number of transmission attempts at each aggregation step in each round; then we find a sub-optimal scheduling policy that improves the convergence rate. Finally, we compare our algorithm with the works that do not consider the transmission success probability and prove its significance.

6.4.1 Effects of Number of Computation and Communication Steps

To study the impact of number of rounds (K^4), number of local update steps during each round (E), and number of transmission attempts at each aggregation step in each round (ℓ), we use a simpler form of **Theorem 6.1**.

Corollary 6.1. *When $\eta_t = \frac{2}{\mu(\gamma+t)}$ with $\gamma = \max\left\{8\frac{L}{\mu}, E\right\}$, after the averaging step at the K-th round, we have*

$$
\mathbb{E}\left[F(w_T) - F^*\right] \leq \frac{L/\mu}{KE}\left[\frac{2}{\mu}\left(\sum_{k=1}^{N}p_k^2\sigma_k^2 + 6L\Gamma + 8(E-1)^2G^2 + 4E^2G^2B\right) + \frac{\mu\gamma}{2}\left\|w_0 - w^*\right\|^2\right],
$$

*where $T = KE$ and B is defined in **Theorem 6.1**.*

---

[4]According to the definition, $K = T/E$ in **Theorem 6.1**.

When $T = KE$ is fixed, the right hand side of **Corollary 6.1** is minimum when $K = T$ and $E = 1$ because we only need to minimize $8(E-1)^2G^2 + 4E^2G^2B$. Therefore, the gap between the global model and the global optimum model is minimized at time T if the server updates the global model after every local update step.

By taking the derivative of the right hand side with respect to E when K (number of FL rounds) is fixed, we observe that the upper bound in **Corollary 6.1** first decreases and then increases; thus, an optimal value for E exists. In fact this can be easily understood by considering that increasing E at first allows each device to move further in the direction of the optimum model parameters; thus, improves the convergence. However, when E is set to a large value, each device moves towards its local optimum model instead of the global optimum.

When E is fixed, it is obvious that the upper bound in **Corollary 6.1** is a decreasing function of K and ℓ, i.e. the the gap between the global model and the global optimum model decreases as we increase the number of communications. However, increasing K decreases the gap more than increasing ℓ. To prove this statement, we denote the success probability after ℓ transmission attempts by $U_k^{(\ell)}$. According to (D.1),

$$
\begin{aligned}
U_k^{(\ell)} &= 1 - \mathbb{E}_{\Phi_\mathrm{I}}\left[\left(1 - e^{-\theta\sigma^2 r_k^\alpha} \prod_{x\in\Phi_\mathrm{I}} \frac{1}{1+\theta r_k^\alpha\|x\|^{-\alpha}} \right)^\ell \right] \\
&\overset{(a)}{\leq} 1 - \mathbb{E}_{\Phi_\mathrm{I}}\left[1 - \ell e^{-\theta\sigma^2 r_k^\alpha} \prod_{x\in\Phi_\mathrm{I}} \frac{1}{1+\theta r_k^\alpha\|x\|^{-\alpha}} \right] \\
&\leq \ell U_k^{(1)},
\end{aligned}
\tag{6.12}
$$

where, in (a), we have used Bernoulli's inequality [113]. When we increase K to ℓK, $\ell > 1$, from the upper bound in **Corollary 6.1** we observe

$$
\begin{aligned}
\frac{L/\mu}{\ell KE}\left[\mathbf{Constant}_1 + \mathbf{Constant}_2 \times \sum_{k=1}^{N} \frac{p_k}{q_k U_k^{(1)}} \right] &\leq \frac{L/\mu}{KE}\left[\mathbf{Constant}_1 + \mathbf{Constant}_2 \times \sum_{k=1}^{N} \frac{p_k}{\ell q_k U_k^{(1)}} \right] \\
&\overset{(a)}{\leq} \frac{L/\mu}{KE}\left[\mathbf{Constant}_1 + \mathbf{Constant}_2 \times \sum_{k=1}^{N} \frac{p_k}{q_k U_k^{(\ell)}} \right],
\end{aligned}
$$

where we have used (6.12) in (a). Although increasing K is more powerful compared to increasing ℓ in decreasing the gap, it consumes more resources since, during each round of the learning process, the scheduled devices perform E steps of local update.

6.4.2 Scheduling Policy

The optimal scheduling policy minimizes the global loss function, and can be obtained from the solution to the following optimization problem:

$$\begin{aligned}
\underset{\{q_k\}}{\text{minimize}} \quad & \mathbb{E}\left[F(w_T)\right] \\
\text{subject to} \quad & \sum_{k=1}^{N} q_k = M, \\
& q_k > 0, \qquad \forall k \in \{1, ..., N\}.
\end{aligned} \qquad (6.13)$$

where the first condition is due to the fact that the BS allocates M resource blocks for the learning process.

To solve (6.13), we need to understand how scheduling affects the global loss for a given w (global model parameters). Since in general this is impossible [30], we use the convergence bound in **Theorem 6.1** to approximately solve (6.13). Thus, we can obtain a sub-optimal solution to (6.13) by solving

$$\begin{aligned}
\underset{\{q_k\}}{\text{minimize}} \quad & \sum_{k=1}^{N} \frac{p_k}{U_k} q_k^{-1} \\
\text{subject to} \quad & \sum_{k=1}^{N} q_k = M, \\
& q_k > 0, \qquad \forall k \in \{1, ..., N\}.
\end{aligned} \qquad (6.14)$$

Now, consider the following optimization

$$\begin{aligned}
\underset{\{q_k\}}{\text{minimize}} \quad & \sum_{k=1}^{N} \frac{p_k}{U_k} q_k^{-1} \\
\text{subject to} \quad & \sum_{k=1}^{N} q_k \leq M, \\
& q_k > 0, \qquad \forall k \in \{1, ..., N\}.
\end{aligned} \qquad (6.15)$$

By contradiction, one can easily prove that solution to (6.15), denoted by $q_1^*, ..., q_N^*$, satisfies $\sum_{k=1}^{N} q_k^* = M$. Thus, (6.15) is equivalent to (6.14), i.e. $q_1^*, ..., q_N^*$ is also the solution to (6.14). (6.15) is a geometric program, and it can be transformed to a convex problem [97]. Therefore, we can find $q_1^*, ..., q_N^*$ using CVX or any other convex solver [114]. Finally, the (sub-) optimal[5] policy for *Scheme II* is achieved when $\hat{q}_k = \frac{q_k^*}{M}$. Note that the optimal policy for *Scheme II* does not necessarily perform better than *Scheme I* since convergence rate and the upper bound in **Theorem 6.1** for *Scheme I* are different from *Scheme II*. In fact, it is only

---

[5]For the rest of the chapter, by optimal scheduling policy we mean the scheduling policy that minimizes the convergence bound in **Theorem 6.1**. This scheduling policy is optimal for (6.14) and sub-optimal for (6.13).

guaranteed that, with optimal policy for *Scheme II*, the upper bound in **Theorem 6.1** performs better compared to other *Scheme II* scheduling policies.

Finally, it is worth mentioning that, according to **Lemma D.2**, the optimal scheduling policy minimizes the variance of the updated global model at each averaging step.

6.4.3 Comparison with the Existing Works

Existing works, such as [35, 36, 37, 38], assume that the BS always successfully recovers the transmitted information of the scheduled devices. For example, [36, 37, 40] use sampling *Scheme II* with probabilities $\{p_k\}$ besides the following averaging approach:

$$w_t = \frac{1}{M} \sum_{k=1}^{N} \sum_{m=1}^{M} \mathbf{1}\left(k \in \mathcal{S}_{t-E}(m)\right) v_t^k \tag{6.16}$$

at $t \in \mathcal{I}_E$. To study this averaging approach when unsuccessful transmission probability is greater than zero, we modify it as

$$w_t = \begin{cases} w_{t-E} \\ \qquad \text{if } \sum_{k=1}^{N} \sum_{m=1}^{M} \mathbf{1}\left(k \in \mathcal{S}_{t-E}(m), \text{SINR}_{k,m} > \theta\right) = 0 \\ \frac{1}{M} \sum_{k=1}^{N} \sum_{m=1}^{M} \mathbf{1}\left(k \in \mathcal{S}_{t-E}(m), \text{SINR}_{k,m} > \theta\right) v_t^k \\ \qquad \text{otherwise} \end{cases} \tag{6.17}$$

In the following, we prove that this averaging scheme does not solve (6.1) when unsuccessful transmission probability is greater than zero and varies across the devices. In this regard, we focus on the following averaging approach:

$$w_t = w_{t-E} + \sum_{k=1}^{N} \sum_{m=1}^{M} \frac{\mathbf{1}\left(k \in \mathcal{S}_{t-E}(m), \text{SINR}_{k,m} > \theta\right)}{\sum_{k=1}^{N} \sum_{m=1}^{M} \mathbf{1}\left(k \in \mathcal{S}_{t-E}(m), \text{SINR}_{k,m} > \theta\right)} \left(v_t^k - w_{t-E}\right), \qquad t \in \mathcal{I}_E, \tag{6.18}$$

where we define $\frac{0}{0} = 0$. It is straightforward to show that, when $M = 1$, (6.18) is same as (6.17).

In **Appendix D.3**, we prove that by using (6.18) for updating the global model parameters at $t \in \mathcal{I}_E$, instead of (6.1), the algorithm converges to the solution to the following problem:

$$\min_w \quad \hat{F}(w) = \sum_{k=1}^{N} p_k U_k F_k(w). \tag{6.19}$$

Since (6.18) and (6.17) are equal for $M = 1$, we can conclude that the employed learning approach by

[36, 37, 40] does not always solve (6.1). In fact, one can easily understand that (6.17) is biased towards devices with higher success probabilities.

Finally, it is worth mentioning that (6.16) can be considered as a special case of (6.6), i.e. (6.6) simplifies to (6.16) when all transmissions are successful ($\text{SINR}_{k,m} > \theta$ and $U_k = 1$, $\forall k, m$).

6.5   Simulation Results and Discussion

Cellular Network: BS intensity λ is 0.001 [points/area]. We assume in each cell there are $N = 100$ user devices and each BS allocates only $M = 20$ resource blocks for the distributed learning process. We also set $\sigma^2 = 10^{-4}$, $\alpha = 4$, and $\theta = -15\,\text{dB}$.

Datasets: We evaluate our proposed federated learning on both real and synthetic datasets. For real data, we use MNIST [115] and distribute our data similar to [37] in a non-i.i.d. fashion. Specifically, each sample in MNIST dataset is a 28×28 image of a handwritten digit between 0 to 9. We distribute the dataset samples such that each device has samples of only two digits and the number of samples at different devices is different. For synthetic data, we follow the same setup as in [116, 37, 36]. In this regard, we generate samples at device k, denoted by (X_k, Y_k), using $y = \text{argmax}\,(\text{softmax}\,(W_k x + b_k))$, where $W_k \in \mathbb{R}^{10\times60}$ and $b_k \in \mathbb{R}^{10}$. Each element in W_k and b_k is a realization of $\mathcal{N}(\mu_k, 1)$, where $\mu_k \sim \mathcal{N}(0, \tilde{\alpha})$. Thus, $\tilde{\alpha}$ controls the degree of difference between local models at different devices[6] [36]. Moreover, $x \in \mathbb{R}^{60}$. The j-th element (feature) in x is drawn from $\mathcal{N}(v_{k,j}, j^{-1.2})$, where $v_{k,j} \sim \mathcal{N}(B_k, 1)$[7] and $B_k \sim \mathcal{N}(0, \tilde{\beta})$. Thus, $\tilde{\beta}$ controls the data heterogeneity. Finally, the number of data samples at device k, denoted by n_k, follows a power law distribution. In this chapter, we set $\tilde{\alpha} = 1$ and $\tilde{\beta} = 1$.

Model: To examine the performance of the proposed FL algorithm on the discussed datasets, we use a three layer neural network with 300 hidden units at each hidden layer. For MNIST datatset, input of the model is a flattened 784-dimensional image. For the synthetic dataset, an input of the model has 60 dimensions. We use ReLU (rectified linear unit) activation function for the hidden layers and softmax in the output layer. At the beginning of round k, $k \in \{0, 1, 2, ..., K - 1\}$, we set the learning rate as $\eta_t = \frac{\eta_0}{1+k}$, where $kE \le t < (k+1)E$. We measure the performance of our model with regularized cross-entropy loss, where we use ℓ_2-norm regularization with regularization parameter 10^{-4}. Therefore, the local loss at device k is computed by

$$F_k(w) = \frac{1}{n_k} \sum_{(x,y)\in\mathcal{D}_k} \text{CrossEntropy}(f(w;x), y) + 10^{-4}\|w\|_2^2,$$

---

[6]When $\tilde{\alpha} = 0$, elements of W_k and b_k at all devices are drawn from the same distribution, i.e. $\mathcal{N}(0, 1)$.

[7]$v_{k,j}$ remains the same across j-th feature of different samples of device k.

where $f(w; x)$ denotes the neural network prediction for input x [37]. The global loss is a weighted average of the local losses as given in (6.1).

Benchmark: When $M = N$ and $U_k = 1$, $\forall k\{1, ..., N\}$, the BS receives the local model parameters of all devices successfully. Under this assumption, if all devices use their full batch and $E = 1$, the distributed learning algorithm is equivalent to the centralized full batch gradient descent. This scenario can be regarded as a benchmark for our proposed algorithm. In Fig. 6.2, performance of the centralized full gradient descent with $\eta_0 = 1$ is shown in terms of global loss and accuracy over both MNIST and synthetic datasets. Accuracy is defined as the percentage of correct predictions.

(a) Global loss. (b) Accuracy.
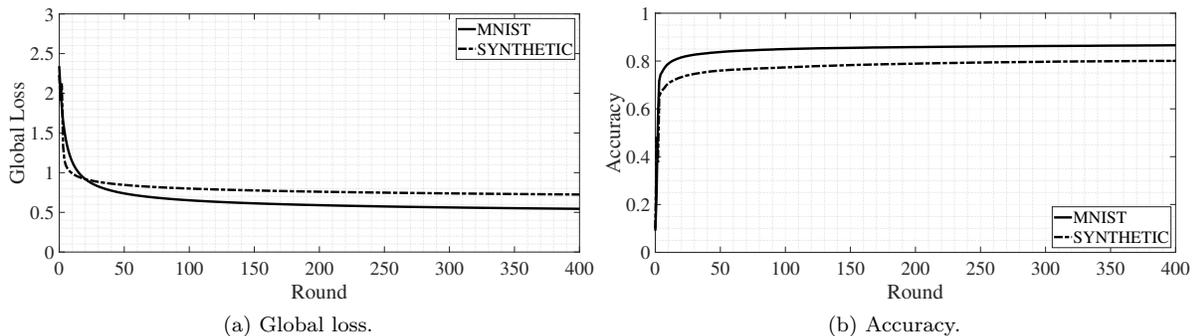
Figure 6.2: Performance of the benchmark algorithm over MNIST and Synthetic datasets.

Training: To train the local models at local SGD steps, for MNIST dataset, we use batch sizes of 64 and set $\eta_0 = 1$. For the synthetic dataset, we use batch sizes of 25 and set $\eta_0 = 0.1$. Since we use a smaller learning rate for the synthetic dataset, we consider higher values for E at the time of simulation. Our codes are available at [117].

Results[8]: In Fig. 6.3(a) and (b), we measure the performance of the proposed FL algorithm in terms of global loss and accuracy on MNIST dataset; Fig. 6.3(c) and (d) also illustrate the performance on the synthetic dataset. For *Scheme II*, we show the results with $\hat{q}_k = \frac{1}{N}$, i.e. uniform selection and $\hat{q}_k = \frac{q_k^*}{M}$, i.e. (sub-) optimal selection. For MNIST dataset, the objective function of (6.15) at $q_1 = q_2 = ... = q_N = \frac{M}{N}$ is 7.09 and at the optimal point $q_1^*, q_2^*, ..., q_N^*$ is 6.48. For the synthetic dataset, these values are 6.89 and 3.66, respectively. Thus, optimal scheduling has more effect on the synthetic dataset which can also be understood from Fig. 6.3. By calculating B in **Theorem 6.1** for *Scheme I* and *Scheme II* with optimal scheduling, we expect that *Scheme I* performs better on MNIST dataset and *Scheme II* performs better on the synthetic dataset. This is also illustrated is Fig. 6.3 and is in compliance with our discussion in **Section 6.4.2**. Moreover, as explained in **Section 6.4.3**, previous works, such as [36, 37, 40], cannot be used for wireless networks since they do not include unsuccessful transmission probability. This is also shown in

---

[8]To make a fair comparison, we have illustrated the average performance over multiple trials.

Fig. 6.3, where we illustrate the performance of the introduced setup in **Section 6.4.3**.

(a) MNIST.

(b) MNIST.

(c) Synthetic.

(d) Synthetic.
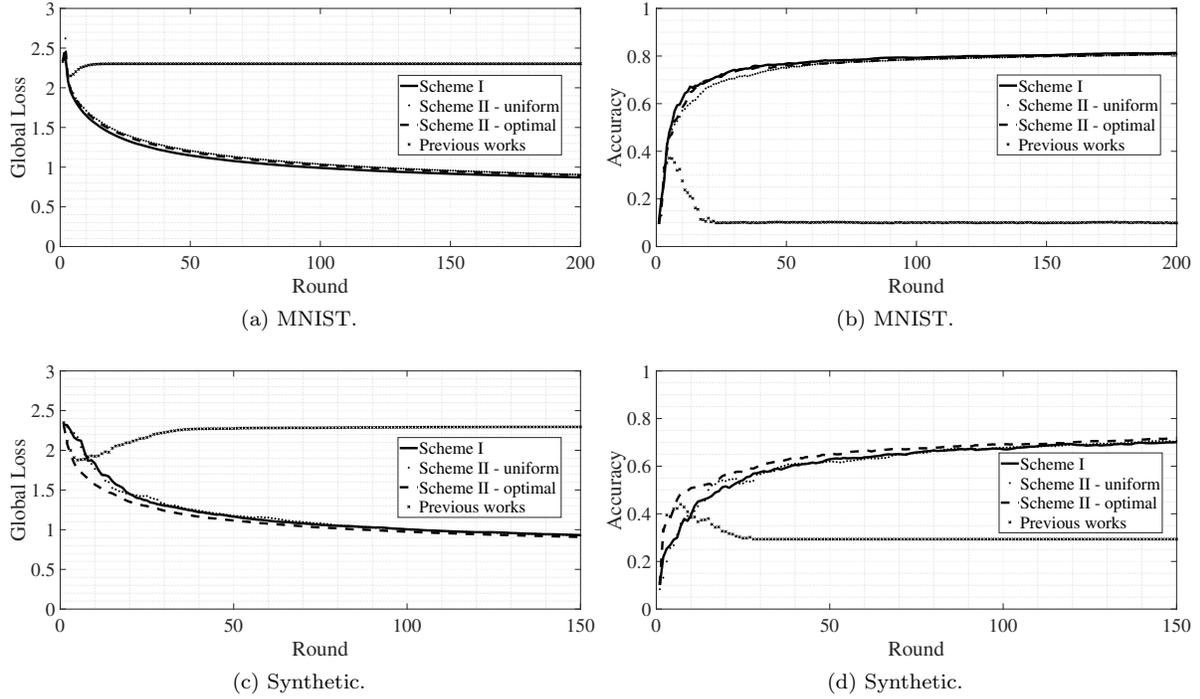
Figure 6.3: In (a) and (b), $\ell = 2$ and $E = 1$. In (c) and (d), $\ell = 2$ and $E = 20$.

We provide more simulation results in Fig. 6.4, 6.5, 6.6, and 6.7 where we study the effects of K, ℓ, and E on the performance of our algorithm. For *Scheme I* sampling, we show the results in Fig. 6.4 and 6.6. For *Scheme II* sampling with $\hat{q}_k = \frac{1}{N}$, we show the results in Fig. 6.5 and 6.7. Note that, when $\hat{q}_k = \frac{1}{N}$ in *Scheme II*, both schemes select users uniformly; however, *Scheme I* selects without replacement, while replacement is allowed in *Scheme II*. We see that *Scheme I* performs better than *Scheme II* with uniform selection. This can be also understood from **Theorem 6.1**. Finally, it is worth mentioning that, with appropriate tuning of the hyper-parameters (K, ℓ, and E), we can reach the benchmarked performance levels.

6.6 Conclusion

Federated learning algorithms cannot be employed in the real world scenarios unless they consider the scarcity of radio resources and unreliability of wireless transmissions. In this regard, in this chapter, we have proposed a federated learning algorithm that is tailored for unreliable and resource-constrained wireless networks, where success probability varies across different devices. Our proposed federated learning algorithm incorporates the success probability in the averaging step. Thus, as the first step towards designing the FL algorithm, we have used stochastic geometry tools to calculate the success probability. We have studied the

(a) Global loss.

(b) Accuracy.

(c) Global loss.

(d) Accuracy.
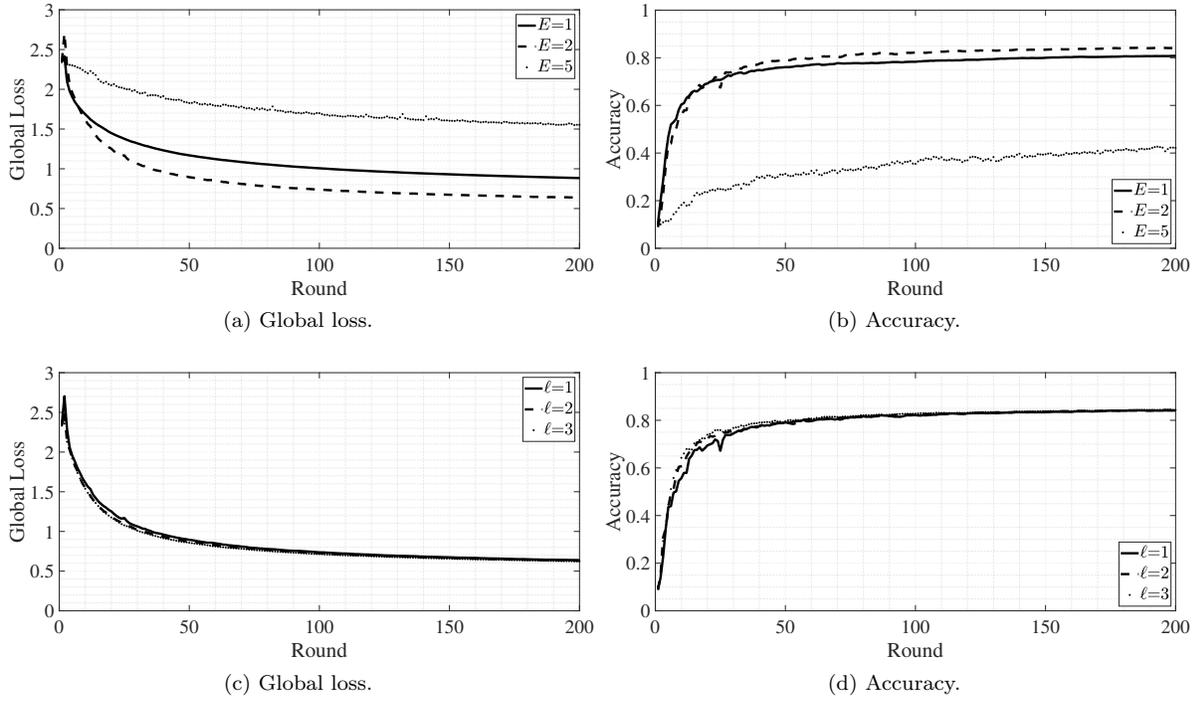
Figure 6.4: Performance of the proposed FL with *Scheme I* sampling over MNIST dataset. In (a) and (b), $\ell = 1$. In (c) and (d), $E = 2$.

(a) Global loss.

(b) Accuracy.

(c) Global loss.

(d) Accuracy.
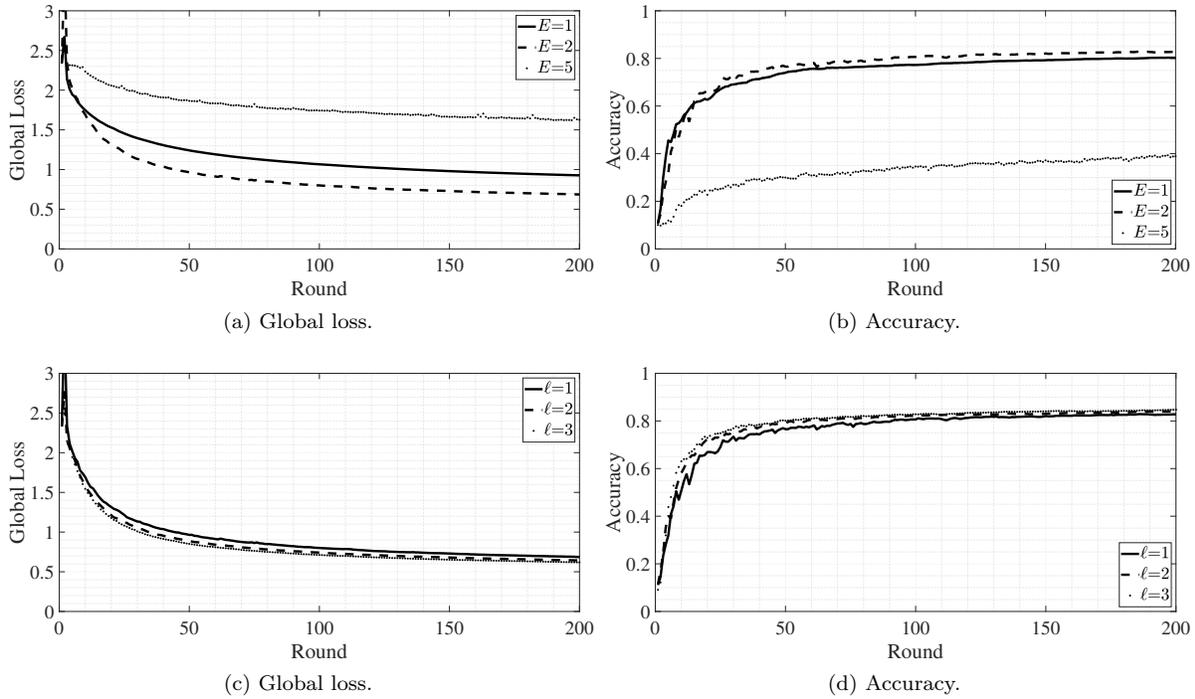
Figure 6.5: Performance of the proposed FL with *Scheme II* sampling with $\hat{q}_k = \frac{1}{N}$, for all k, over MNIST dataset. In (a) and (b), $\ell = 1$. In (c) and (d), $E = 2$.

(a) Global loss.

(b) Accuracy.

(c) Global loss.

(d) Accuracy.
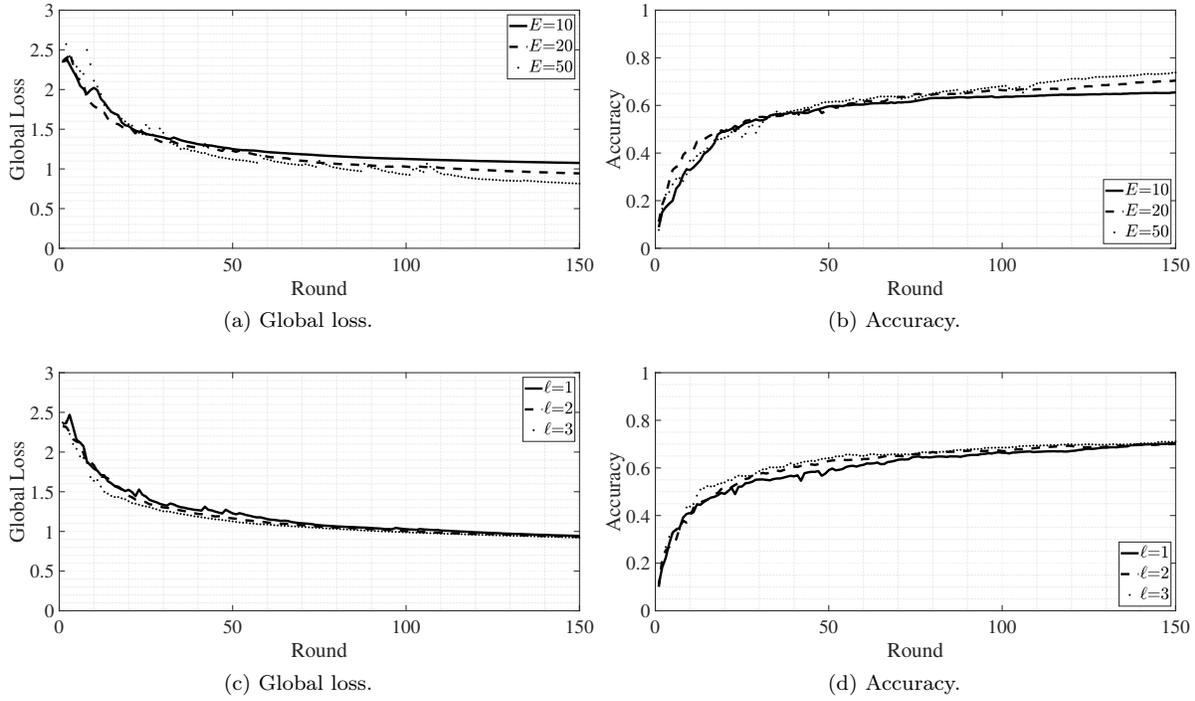
Figure 6.6: Performance of the proposed FL with *Scheme I* sampling over the synthetic dataset. In (a) and (b), $\ell = 1$. In (c) and (d), $E = 20$.

(a) Global loss.

(b) Accuracy.

(c) Global loss.

(d) Accuracy.
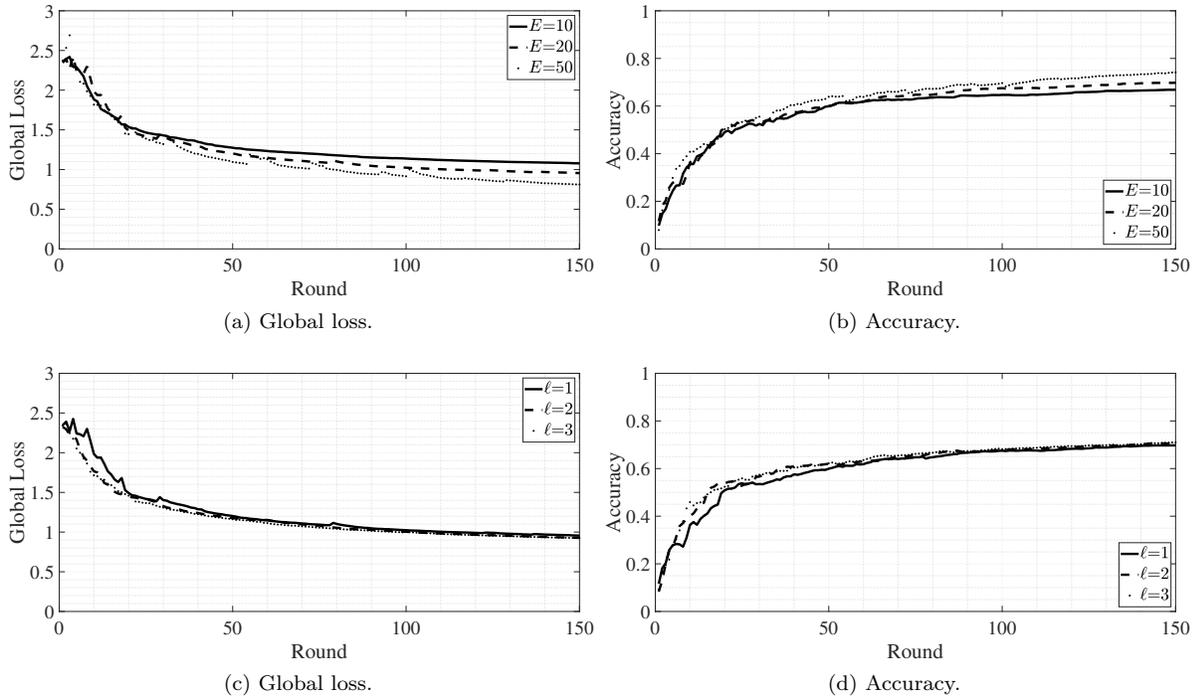
Figure 6.7: Performance of the proposed FL with *Scheme II* sampling with $\hat{q}_k = \frac{1}{N}$, for all k, over the synthetic dataset. In (a) and (b), $\ell = 1$. In (c) and (d), $E = 20$.

convergence of the proposed algorithm. Specifically, we have proven that the algorithm converges with rate $\mathcal{O}(\frac{1}{T})$ for strongly convex and smooth problems on non-i.i.d. data. The effects of computation, communication, and scheduling on the convergence rate of the algorithm have also been investigated. Finally, we have verified our algorithm through experimenting on real and synthetic datasets.

Chapter 7

Summary and Future Directions

7.1 Summary

In Chapter 3, using tools from stochastic geometry, we have derived the conditional success probability (CSP), its moments, and its CCDF (meta distribution) for uplink and downlink NOMA. The spatial distribution of uplink interferers is required for deriving the analytical results. We have proposed two point process models for spatial locations of the uplink inter-cell interferers by moment matching (BS/user pair correlation function matching). Using these models, in Chapter 4, we removed some simplifying assumptions that we make in Chapter 3. Specifically, we derived the uplink coverage probability (first moment of CSP) while including interference correlation and removing distance-based ranking assumption.

Most of the existing works assume that ranking users in each NOMA cluster based on their distances, instead of the complete CSI, is a valid approximation. In Chapter 4, we studied the effects of this approximation on the coverage probability analysis in the uplink as well as in the downlink. Specifically, the accuracy probability, which is the probability that the distance-based ranking matches ranking based on the instantaneous signal power, has been defined and derived for Rayleigh and Nakagami-m fading channels. Effects of NOMA cluster size, path-loss, and user selection on the accuracy probability are also investigated.

In Chapter 5, we have derived the distribution and mean of the sojourn time in multi-tier cellular networks. In multi-tier networks with maximum biased averaged received power association we need the chord length distribution in weighted Poisson Voronoi tessellation. In Chapter 5, we have derived the area of contact in weighted Poisson Voronoi tessellation from which we have obtained the linear contact distribution function. Then, chord length distribution is obtained from linear contact distribution function. Chapter 5 also studies the relation between mean sojourn time and other mobility-related performance metrics. Specifically, it

shows that mean sojourn time is inversely proportional to the handoff rate. Effects of different parameters on mean and distribution of sojourn time are also investigated in Chapter 5.

In Chapter 6, we study the performance of distributed machine learning algorithms in cellular wireless networks. Previous works ignored the scarcity of radio resources and unreliability of wireless transmissions, when designing the learning algorithms. In this regard, in Chapter 6, we have proposed a federated learning algorithm that is tailored for unreliable and resource-constrained wireless networks, where success probability varies across different devices. We have studied the convergence of our algorithm and proved that the algorithm converges with rate $\mathcal{O}(\frac{1}{T})$ for strongly convex and smooth problems on non-i.i.d. data. The effects of computation, communication, and scheduling on the convergence rate of the algorithm have also been investigated in Chapter 6.

7.2   Future Directions

Design and analysis of improved multiple access methods: Meeting performance targets and supporting standard services of 5G and B5G necessitate efficient user access mechanisms. In this dissertation, we have focused on the large-scale performance of power-domain NOMA in single-antenna setup and calculated its gain over orthogonal multiple access (OMA). Although, NOMA seems promising for this scenario, it has been shown recently in [118] that power-domain NOMA is inefficient in multiple-input-multiple-output setups compared to conventional Multi-User Linear Precoding (MU-LP) and Rate-Splitting Multiple Access (RSMA). Moreover, extra overhead, user grouping, receiver design, and complexity of NOMA diminish the gain of NOMA over OMA. Future works need to address the existing challenges of NOMA to make it beneficial for future use-cases. Also, Large-scale performance analysis of RSMA in MIMO setup and comparison with MU-LP and NOMA is an open problem.

Mobility analysis of integrated terrestrial-aerial networks: With stringent requirements in future generations of wireless networks and emerging applications in new vertical sectors, mobility related challenges become more noticeable. In this dissertation, we have studied the sojourn time distribution and its relation to handoff rate for multi-tier terrestrial networks. We can use sojourn time distribution and handoff rate for resource allocation and mobility management to improve network performance. For example, when a user enters a new cell, sojourn time distribution can be employed to avoid an unnecessary handoff based on the user's distance between to the BS. Moreover, analysis of different handoff types and handoff in more complex scenarios is necessary for future applications. For instance, Xn-based and N2-based handoff rate analysis for understanding effects of mobility in 5G networks is required for better resource allocation, or considering

Poisson Line process to model cars mobility in 5G networks can be insightful.

Improved federated learning methods and theorem analyses in wireless networks: Our proposed algorithm modifies FedAvg such that it converges in wireless medium under resource scarcity and transmission unreliability. Improving the convergence rate of the proposed algorithm by predicting clients' drifts, sharing more local information besides the updated local model, and client sampling based on data distribution seem promising. Moreover, FedAvg on non-i.i.d. data lacks theoretical guarantee in a convex optimization setting. Similarly, convergence analysis of our algorithm, which can be regarded as a modified version of FedAvg, with convex loss functions is an open problem. Finally, our results can be used in network design for edge computing and spectrum management [119].

Bibliography

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2020.

[2] "5G new radio: Introduction to the physical layer," White Paper, National Instruments, 2018, [Online]. Available:ni.com/5g/nr.

[3] I. Group *et al.*, "Ieee 5G and beyond technology roadmap white paper," IEEE, Tech. Rep, Tech. Rep., 2017.

[4] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE Journal on Selected Areas in Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.

[5] E. Hossain and M. Hasan, "5G cellular: key enabling technologies and research challenges," *IEEE Instrumentation Measurement Magazine*, vol. 18, no. 3, pp. 11–21, 2015.

[6] M. Vaezi, Z. Ding, and H. V. Poor, *Multiple access techniques for 5G wireless networks and beyond.* Springer, 2019.

[7] A. E. Mostafa, Y. Zhou, and V. W. S. Wong, "Connectivity maximization for narrowband iot systems with noma," in *2017 IEEE International Conference on Commun. (ICC)*, May 2017.

[8] H. Tabassum, M. S. Ali, E. Hossain, M. J. Hossain, and D. I. Kim, "Uplink vs. downlink NOMA in cellular networks: Challenges and research directions," in *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, June 2017.

[9] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, June 2013.

[10] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Processing Letters*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[11] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G non-orthogonal multiple-access downlink transmissions," *IEEE Trans. on Vehicular Technology*, vol. 65, no. 8, pp. 6010–6023, 2016.

[12] N. Zhang, J. Wang, G. Kang, and Y. Liu, "Uplink non-orthogonal multiple access in 5G systems," *IEEE Commun. Letters*, vol. 20, no. 3, pp. 458–461, Mar. 2016.

[13] T. Takeda and K. Higuchi, " Enhanced user fairness using non-orthogonal access with SIC in cellular uplink," *in Proc. of IEEE Vehicular Technology Conference (VTC Fall)*, Sep. 2011.

[14] Y. Endo, Y. Kishiyama, and K. Higuchi, "Uplink non-orthogonal access with MMSE-SIC in the presence of inter-cell interference," *in Proc. of IEEE ISWCS*, 2012.

[15] 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 Trans. on Commun.*, vol. 65, no. 8, pp. 3555–3570, Aug 2017.

[16] Z. Zhang, H. Sun, R. Q. Hu, and Y. Qian, "Stochastic geometry based performance study on 5G non-orthogonal multiple access scheme," *IEEE Global Commun. Conference (GLOBECOM'16)*, Dec. 2016.

[17] 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 Commun.*, vol. 35, no. 12, pp. 2771–2784, 2017.

[18] Y. Liu, Z. Qin, M. Elkashlan, A. Nallanathan, and J. A. McCann, "Non-orthogonal multiple access in large-scale heterogeneous networks," *IEEE Journal on Selected Areas in Commun.*, vol. 35, no. 12, pp. 2667–2680, 2017.

[19] C. Liu and D.-C. Liang, "Heterogeneous networks with power-domain NOMA: Coverage, throughput, and power allocation analysis," *IEEE Trans. on Wireless Commun.*, vol. 17, no. 5, pp. 3524–3539, 2018.

[20] Y. Wang, M. Haenggi, and Z. Tan, "The meta distribution of the SIR for cellular networks with power control," *IEEE Trans. on Commun.*, vol. 66, no. 4, pp. 1745–1757, Apr. 2018.

[21] J. G. Andrews, A. K. Gupta, and H. S. Dhillon, "A primer on cellular network analysis using stochastic geometry," *arXiv preprint arXiv:1604.03183*, 2016.

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

[23] H. Tabassum, M. Salehi, and E. Hossain, "Fundamentals of mobility-aware performance characterization of cellular networks: A tutorial," *IEEE Commun. Surveys & Tutorials*, vol. 21, no. 3, pp. 2288–2308, thirdquarter 2019.

[24] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on uavs for wireless networks: Applications, challenges, and open problems," *IEEE Commun. Surveys & Tutorials*, vol. 21, no. 3, pp. 2334–2360, 2019.

[25] A. L. E. Corral-Ruiz, F. A. Cruz-Pérez, and G. Hernandez-Valdez, "Channel holding time in mobile cellular networks with heavy-tailed distributed cell dwell time," in *2011 IEEE Wireless Commun. and Networking Conference*, March 2011, pp. 1242–1247.

[26] C. Lee and Z. Syu, "Handover analysis of macro-assisted small cell networks," in *2014 IEEE International Conference on Internet of Things (iThings), and IEEE Green Computing and Commun. (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom)*, Sep. 2014, pp. 604–609.

[27] X. Xu, Z. Sun, X. Dai, T. Svensson, and X. Tao, "Modeling and analyzing the cross-tier handover in heterogeneous networks," *IEEE Trans. on Wireless Commun.*, vol. 16, no. 12, pp. 7859–7869, Dec 2017.

[28] Y. Hong, X. Xu, M. Tao, J. Li, and T. Svensson, "Cross-tier handover analyses in small cell networks: A stochastic geometry approach," in *2015 IEEE International Conference on Commun. (ICC)*, June 2015, pp. 3429–3434.

[29] X. Lin, R. K. Ganti, P. J. Fleming, and J. G. Andrews, "Towards understanding the fundamentals of mobility in cellular networks," *IEEE Trans. on Wireless Commun.*, vol. 12, no. 4, pp. 1686–1698, April 2013.

[30] S. Wang, T. Tuor, T. Salonidis, K. K. Leung, C. Makaya, T. He, and K. Chan, "Adaptive federated learning in resource constrained edge computing systems," *IEEE Journal on Selected Areas in Commun.*, vol. 37, no. 6, pp. 1205–1221, 2019.

[31] C. Dinh, N. H. Tran, M. N. Nguyen, C. S. Hong, W. Bao, A. Zomaya, and V. Gramoli, "Federated learning over wireless networks: Convergence analysis and resource allocation," *arXiv preprint arXiv:1910.13067*, 2019.

[32] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.

[33] L. U. Khan, W. Saad, Z. Han, E. Hossain, and C. S. Hong, "Federated learning for internet of things: Recent advances, taxonomy, and open challenges," *arXiv preprint arXiv:2009.13012*, 2020.

[34] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings *et al.*, "Advances and open problems in federated learning," *arXiv preprint arXiv:1912.04977*, 2019.

[35] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics*. PMLR, 2017, pp. 1273–1282.

[36] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proceedings of Machine Learning and Systems*, vol. 2, pp. 429–450, 2020.

[37] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, "On the convergence of fedavg on non-iid data," *arXiv preprint arXiv:1907.02189*, 2019.

[38] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "Scaffold: Stochastic controlled averaging for federated learning," in *International Conference on Machine Learning*. PMLR, 2020, pp. 5132–5143.

[39] F. Haddadpour and M. Mahdavi, "On the convergence of local descent methods in federated learning," *arXiv preprint arXiv:1910.14425*, 2019.

[40] H. T. Nguyen, V. Sehwag, S. Hosseinalipour, C. G. Brinton, M. Chiang, and H. V. Poor, "Fast-convergent federated learning," *IEEE Journal on Selected Areas in Commun.*, 2020.

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

[42] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications*. John Wiley & Sons, 2013.

[43] M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, "Stochastic geometry and random graphs for the analysis and design of wireless networks," *IEEE journal on selected areas in commun.*, vol. 27, no. 7, pp. 1029–1046, 2009.

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

[45] C. Saha, M. Afshang, and H. S. Dhillon, "3gpp-inspired hetnet model using poisson cluster process: Sum-product functionals and downlink coverage," *IEEE Trans. on Commun.*, vol. 66, no. 5, pp. 2219–2234, 2017.

[46] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey," *IEEE Commun. Surveys & Tutorials*, vol. 15, no. 3, pp. 996–1019, 2013.

[47] H. ElSawy, A. Sultan-Salem, M.-S. Alouini, and M. Z. Win, "Modeling and analysis of cellular networks using stochastic geometry: A tutorial," *IEEE Commun. Surveys & Tutorials*, vol. 19, no. 1, pp. 167–203, 2016.

[48] R. K. Ganti and M. Haenggi, "Asymptotics and approximation of the sir distribution in general cellular networks," *IEEE Trans. on Wireless Commun.*, vol. 15, no. 3, pp. 2130–2143, 2015.

[49] M. Haenggi, "The meta distribution of the SIR in poisson bipolar and cellular networks," *IEEE Trans. on Wireless Commun.*, vol. 15, pp. 2577–2589, Apr. 2016.

[50] J. Gil-Pelaez, "Note on the inversion theorem," *Biometrika*, vol. 38, no. 3-4, pp. 481–482, 1951.

[51] Q. Cui, X. Yu, Y. Wang, and M. Haenggi, "The SIR meta distribution in poisson cellular networks with base station cooperation," *IEEE Trans. on Commun.*, vol. 66, no. 3, pp. 1234–1249, 2018.

[52] M. Salehi, A. Mohammadi, and M. Haenggi, "Analysis of D2D underlaid cellular networks: SIR meta distribution and mean local delay," *IEEE Trans. on Commun.*, vol. 65, pp. 2904–2916, July 2017.

[53] N. Deng and M. Haenggi, "A fine-grained analysis of millimeter-wave device-to-device networks," *IEEE Trans. on Commun.*, vol. 65, no. 11, pp. 4940–4954, Nov. 2017.

[54] F. Baccelli and B. Blaszczyszyn, "A new phase transitions for local delays in MANETs," in *Proc. of IEEE Infocom'10*, Mar. 2010.

[55] M. Haenggi, "The local delay in poisson networks," *IEEE Trans. on Information Theory*, vol. 59, pp. 1788–1802, Mar. 2013.

[56] E. Hossain, M. Rasti, H. Tabassum, and A. Abdelnasser, "Evolution toward 5G multi-tier cellular wireless networks: An interference management perspective," *IEEE Wireless Commun.*, vol. 21, no. 3, pp. 118–127, June 2014.

[57] M. Bennis, M. Debbah, and H. V. Poor, "Ultrareliable and low-latency wireless communication: Tail, risk, and scale," *Proceedings of the IEEE*, vol. 106, no. 10, pp. 1834–1853, 2018.

[58] J. G. Andrews, R. K. Ganti, M. Haenggi, N. Jindal, and S. Weber, "A primer on spatial modeling and analysis in wireless networks," *IEEE Commun. Magazine*, vol. 48, no. 11, pp. 156–163, Nov. 2010.

[59] G. Geraci, M. Wildemeersch, and T. Q. S. Quek, "Energy efficiency of distributed signal processing in wireless networks: A cross-layer analysis," *IEEE Trans. on Signal Processing*, vol. 64, no. 4, pp. 1034–1047, Feb. 2016.

[60] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, and Z. Wang, "Non-orthogonal multiple access for 5g: solutions, challenges, opportunities, and future research trends," *IEEE Commun. Magazine*, vol. 53, no. 9, pp. 74–81, 2015.

[61] Z. Wei, J. Yuan, D. W. K. Ng, M. Elkashlan, and Z. Ding, "A survey of downlink non-orthogonal multiple access for 5g wireless communication networks," *arXiv preprint arXiv:1609.01856*, 2016.

[62] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5g networks: Research challenges and future trends," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2181–2195, 2017.

[63] H. David and H. Nagaraja, *Order Statistics*, 3rd ed.   Wiley, New York, NY, 2003.

[64] H.-C. Yang and M.-S. Alouini, *Order Statistics in Wireless Communications: Diversity, Adaptation, and Scheduling in MIMO and OFDM Systems*.   Cambridge University Press, 2011.

[65] M. Afshang, H. S. Dhillon, and P. H. J. Chong, "Modeling and performance analysis of clustered device-to-device networks," *IEEE Trans. on Wireless Commun.*, vol. 15, no. 7, pp. 4957–4972, July 2016.

[66] M. A. Sedaghat and R. R. Müller, "On user pairing in uplink NOMA," *IEEE Trans. on Wireless Commun.*, vol. 17, no. 5, pp. 3474–3486, May 2018.

[67] H. Xing, Y. Liu, A. Nallanathan, Z. Ding, and H. V. Poor, "Optimal throughput fairness tradeoffs for downlink non-orthogonal multiple access over fading channels," *IEEE Trans. on Wireless Commun.*, vol. 17, no. 6, pp. 3556–3571, June 2018.

[68] Z. Zhang, Z. Ma, M. Xiao, G. Liu, and P. Fan, "Modeling and analysis of non-orthogonal MBMS transmission in heterogeneous networks," *IEEE Journal on Selected Areas in Commun.*, vol. 35, no. 10, pp. 2221–2237, Oct. 2017.

[69] J. Wang, B. Xia, K. Xiao, Y. Gao, and S. Ma, "Outage performance analysis for wireless non-orthogonal multiple access systems," *IEEE Access*, vol. 6, pp. 3611–3618, 2018.

[70] X. Chen, Z. Zhang, C. Zhong, D. W. K. Ng, and R. Jia, "Exploiting inter-user interference for secure massive non-orthogonal multiple access," *IEEE Journal on Selected Areas in Commun.*, vol. 36, no. 4, pp. 788–801, 2018.

[71] M. Haenggi, "On distances in uniformly random networks," *IEEE Trans. on Information Theory*, vol. 51, pp. 3584–3586, Oct. 2005.

[72] K. S. Ali, M. Haenggi, H. ElSawy, A. Chaaban, and M.-S. Alouini, "Downlink non-orthogonal multiple access (NOMA) in poisson networks," *IEEE Trans. on Commun.*, vol. 67, no. 2, pp. 1613–1628, 2018.

[73] M. Wildemeersch, T. Q. S. Quek, M. Kountouris, A. Rabbachin, and C. H. Slump, "Successive interference cancellation in heterogeneous networks," *IEEE Trans. on Commun.*, vol. 62, no. 12, pp. 4440–4453, Dec. 2014.

[74] Z. Shi, S. Ma, H. ElSawy, G. Yang, and M.-S. Alouini, "Cooperative HARQ-assisted NOMA scheme in large-scale D2D networks," *IEEE Trans. on Commun.*, vol. 66, no. 9, pp. 4286–4302, 2018.

[75] J. Choi, "On HARQ-IR for downlink NOMA systems," *IEEE Trans. on Commun.*, vol. 64, no. 8, pp. 3576–3584, Aug. 2016.

[76] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE Journal on Selected Areas in Commun.*, vol. 34, no. 4, pp. 938–953, Apr. 2016.

[77] Z. Ding, R. Schober, and H. V. Poor, "A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment," *IEEE Trans. on Wireless Commun.*, vol. 15, no. 6, pp. 4438–4454, June 2016.

[78] H. Chingoska, Z. Hadzi-Velkov, I. Nikoloska, and N. Zlatanov, "Resource allocation in wireless powered communication networks with non-orthogonal multiple access," *IEEE Wireless Commun. Letters*, vol. 5, no. 6, pp. 684–687, Dec 2016.

[79] S. M. Yu and S. 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)*, May 2013, pp. 119–124.

[80] M. D. Renzo, A. Zappone, T. T. Lam, and M. Debbah, "System-level modeling and optimization of the energy efficiency in cellular networks – a stochastic geometry framework," *IEEE Trans. on Wireless Commun.*, vol. 17, no. 4, pp. 2539–2556, Apr. 2018.

[81] B. Yu, S. Mukherjee, H. Ishii, and L. Yang, "Dynamic TDD support in the LTE-B enhanced local area architecture," in *2012 IEEE Globecom Workshops*, Dec. 2012, pp. 585–591.

[82] M. Ahsanullah, V. B. Nevzorov, and M. Shakil, *An Introduction to Order Statistics*. Springer, 2013.

[83] A. H. Stroud and D. Secrest, "Gaussian quadrature formulas," 1966.

[84] J. Ma, V. Rokhlin, and S. Wandzura, "Generalized gaussian quadrature rules for systems of arbitrary functions," *SIAM Journal on Numerical Analysis*, vol. 33, no. 3, pp. 971–996, 1996.

[85] E. Hossain, L. Le, and D. Niyato, *Radio Resource Management in Multi-Tier Cellular Wireless Networks*. Wiley, 2013.

[86] S. Shin, U. Lee, F. Dressler, and H. Yoon, "Analysis of cell sojourn time in heterogeneous networks with small cells," *IEEE Commun. Letters*, vol. 20, no. 4, pp. 788–791, April 2016.

[87] W. Bao and B. Liang, "Stochastic geometric analysis of user mobility in heterogeneous wireless networks," *IEEE Journal on Selected Areas in Commun.*, vol. 33, no. 10, pp. 2212–2225, 2015.

[88] S. Sadr and R. S. Adve, "Handoff rate and coverage analysis in multi-tier heterogeneous networks," *IEEE Trans. on Wireless Commun.*, vol. 14, no. 5, pp. 2626–2638, May 2015.

[89] S. Hsueh and K. Liu, "An equivalent analysis for handoff probability in heterogeneous cellular networks," *IEEE Commun. Letters*, vol. 21, no. 6, pp. 1405–1408, June 2017.

[90] H. Fu, P. Lin, and Y. Lin, "Reducing signaling overhead for femtocell/macrocell networks," *IEEE Trans. on Mobile Computing*, vol. 12, no. 8, pp. 1587–1597, Aug 2013.

[91] R. Arshad, H. Elsawy, S. Sorour, T. Y. Al-Naffouri, and M. Alouini, "Handover management in 5g and beyond: A topology aware skipping approach," *IEEE Access*, vol. 4, pp. 9073–9081, 2016.

[92] Y. Fang, I. Chlamtac, and Yi-Bing Lin, "Call performance for a pcs network," *IEEE Journal on Selected Areas in Commun.*, vol. 15, no. 8, pp. 1568–1581, Oct 1997.

[93] F. Khan and D. Zeghlache, "Effect of cell residence time distribution on the performance of cellular mobile networks," in *1997 IEEE 47th Vehicular Technology Conference. Technology in Motion*, vol. 2, May 1997, pp. 949–953 vol.2.

[94] S. Thajchayapong and J. M. Peha, "Mobility patterns in microcellular wireless networks," *IEEE Trans. on Mobile Computing*, vol. 5, no. 1, pp. 52–63, Jan 2006.

[95] A. Merwaday, I. Güvenç, W. Saad, A. Mehbodniya, and F. Adachi, "Sojourn time-based velocity estimation in small cell poisson networks," *IEEE Commun. Letters*, vol. 20, no. 2, pp. 340–343, Feb 2016.

[96] J. Moller, *Lectures on Random Voronoi Tessellations.* Springer Science & Business Media, 2012, vol. 87.

[97] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, 2004.

[98] H.-S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous cellular networks with flexible cell association: A comprehensive downlink sinr analysis," *IEEE Trans. on Wireless Commun.*, vol. 11, no. 10, pp. 3484–3495, 2012.

[99] L. Heinrich, "Contact and chord length distribution of a stationary voronoi tessellation," *Advances in Applied Probability*, vol. 30, no. 3, pp. 603–618, 1998.

[100] A. Hard, K. Rao, R. Mathews, S. Ramaswamy, F. Beaufays, S. Augenstein, H. Eichner, C. Kiddon, and D. Ramage, "Federated learning for mobile keyboard prediction," *arXiv preprint arXiv:1811.03604*, 2018.

[101] W. Y. B. Lim, N. C. Luong, D. T. Hoang, Y. Jiao, Y.-C. Liang, Q. Yang, D. Niyato, and C. Miao, "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Commun. Surveys & Tutorials*, vol. 22, no. 3, pp. 2031–2063, 2020.

[102] M. M. Amiri and D. Gündüz, "Machine learning at the wireless edge: Distributed stochastic gradient descent over-the-air," *IEEE Trans. on Signal Processing*, vol. 68, pp. 2155–2169, 2020.

[103] F. Zhou and G. Cong, "On the convergence properties of a k-step averaging stochastic gradient descent algorithm for nonconvex optimization," *arXiv preprint arXiv:1708.01012*, 2017.

[104] H. H. Yang, Z. Liu, T. Q. Quek, and H. V. Poor, "Scheduling policies for federated learning in wireless networks," *IEEE Trans. on Commun.*, vol. 68, no. 1, pp. 317–333, 2019.

[105] V. Smith, S. Forte, C. Ma, M. Takáč, M. I. Jordan, and M. Jaggi, "Cocoa: A general framework for communication-efficient distributed optimization," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 8590–8638, 2017.

[106] R. Tanbourgi, H. S. Dhillon, J. G. Andrews, and F. K. Jondral, "Effect of spatial interference correlation on the performance of maximum ratio combining," *IEEE Transactions on Wireless Commun.*, vol. 13, no. 6, pp. 3307–3316, 2014.

[107] H. ElSawy and E. Hossain, "On stochastic geometry modeling of cellular uplink transmission with truncated channel inversion power control," *IEEE Trans. on Wireless Commun.*, vol. 13, no. 8, pp. 4454–4469, 2014.

[108] M. Di Renzo and P. Guan, "Stochastic geometry modeling and system-level analysis of uplink heterogeneous cellular networks with multi-antenna base stations," *IEEE Trans. on Commun.*, vol. 64, no. 6, pp. 2453–2476, 2016.

[109] T. Bai and R. W. Heath, "Analyzing uplink sinr and rate in massive mimo systems using stochastic geometry," *IEEE Trans. on commun.*, vol. 64, no. 11, pp. 4592–4606, 2016.

[110] S. Singh, M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "Tractable model for rate in self-backhauled millimeter wave cellular networks," *IEEE Journal on Selected Areas in Commun.*, vol. 33, no. 10, pp. 2196–2211, 2015.

[111] Y. Zhang, J. C. Duchi, and M. J. Wainwright, "Communication-efficient algorithms for statistical optimization," *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 3321–3363, 2013.

[112] S. U. Stich, "Local sgd converges fast and communicates little," *arXiv preprint arXiv:1805.09767*, 2018.

[113] D. S. Mitrinovic and P. M. Vasic, *Analytic inequalities*. Springer, 1970, vol. 61.

[114] M. Grant and S. Boyd, "Cvx: Matlab software for disciplined convex programming, version 2.1," 2014.

[115] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[116] O. Shamir, N. Srebro, and T. Zhang, "Communication-efficient distributed optimization using an approximate newton-type method," in *International Conference on Machine Learning*, 2014, pp. 1000–1008.

[117] M. Salehi and E. Hossain, "Federated learning in unreliable and resource-constrained cellular wireless networks," https://github.com/mhdslh/ Federated-Learning-in-Unreliable-and-Resource-Constrained-Cellular-Wireless-Networks, 2020.

[118] B. Clerckx, Y. Mao, R. Schober, E. Jorswieck, D. J. Love, J. Yuan, L. Hanzo, G. Y. Li, E. G. Larsson, and G. Caire, "Is noma efficient in multi-antenna networks? a critical look at next generation multiple access techniques," *arXiv preprint arXiv:2101.04802*, 2021.

[119] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications: Motivation, opportunities, and challenges," *IEEE Commun. Magazine*, vol. 58, no. 6, pp. 46–51, 2020.

Appendix A

Appendix to Chapter 3

A.1 Proof of Theorem 3.1

We first derive the CSP $P_{\text{s},(m)}$ for the m-th rank uplink NOMA user as follows:

$$
\begin{aligned}
P_{\text{s},(m)}(\theta) &= \mathbb{P}\left(\text{SIR}_{(m)} > \theta \mid \Phi_{\text{U}}, \text{tx}\right) \\
&= \mathbb{P}\left(h_{x_{(m)}} > \theta\|x_{(m)}\|^\alpha \left(\sum_{x \in \Phi_{\text{I}}} h_x \|x\|^{-\alpha} + \sum_{i=m+1}^{N} h_{x_{(i)}} \|x_{(i)}\|^{-\alpha}\right) \mid \Phi_{\text{U}}, \text{tx}\right) \\
&\overset{(a)}{=} \prod_{x \in \Phi_{\text{I}}} \frac{1}{1 + \theta\|x_{(m)}\|^\alpha \|x\|^{-\alpha}} \prod_{i=m+1}^{N} \frac{1}{1 + \theta\|x_{(m)}\|^\alpha \|x_{(i)}\|^{-\alpha}},
\end{aligned}
\tag{A.1}
$$

where (a) follows from applying the CCDF of the unit mean exponential distribution of $h_{x_{(m)}}$ and then the Laplace transform of the unit mean exponential distribution of h_x and $h_{x_{(i)}}$. Note that Φ_{U} represents the superposition of two independent point processes, namely, the inter-cell interferer point process Φ_{I} and point process of users located in the typical Voronoi cell (intra-cell users).

Next, we derive the b-th moment of CSP $M_{b,(m)} = \mathbb{E}_{\Phi_{\text{U}}}\left[P_{\text{s},(m)}^b\right]$ as follows:

$$
M_{b,(m)} \overset{(a)}{=} \mathbb{E}_{x_{(m)}}\left[\underbrace{\mathbb{E}_{\Phi_{\text{I}}}\left[\prod_{x \in \Phi_{\text{I}}}\left(\frac{1}{1 + \theta\|x_{(m)}\|^\alpha \|x\|^{-\alpha}}\right)^b\right]}_{\text{Part A}} \underbrace{\mathbb{E}_r\left[\left(\frac{1}{1 + \theta\|x_{(m)}\|^\alpha r^{-\alpha}}\right)^b\right]^{N-m}}_{\text{Part B}}\right],
\tag{A.2}
$$

where (a) is obtained by noting that (i) Φ_{U} is the superposition of the inter-cell and intra-cell point processes, (ii) the inter-cell interferers' point process and the intra-cell interferers' point process are independent, and (iii) conditioned on the user at rank m, distributions of the distances of the intra-cell interfering users from

118

the typical BS are i.i.d, so we can replace $\|x_{(i)}\|$ with r [15]. Using the **Model 1** for inter-cell interferer point process in Section 3.3, we can approximate **Part A** as follows:

$$\mathbb{E}_{\Phi_{\mathrm{I}}} \left[\prod_{x \in \Phi_{\mathrm{I}}} \left(\frac{1}{1 + \theta\|x_{(m)}\|^{\alpha}\|x\|^{-\alpha}} \right)^{b} \right] \approx \mathbb{E} \left[\prod_{x \in \bar{\Phi}_{\mathrm{P}}} \left(\frac{1}{1 + \theta\|x_{(m)}\|^{\alpha}\|x\|^{-\alpha}} \right)^{Nb} \mid x_{(m)} \right]$$

$$\overset{(a)}{=} \exp \left(-2\pi\lambda_{\mathrm{b}} \int_{0}^{\infty} \left[1 - \left(\frac{1}{1 + \theta\|x_{(m)}\|^{\alpha}r^{-\alpha}} \right)^{Nb} \right] \left(1 - e^{-(12/5)\lambda_{\mathrm{b}}\pi r^{2}} \right) r\mathrm{d}r \right),$$

where we approximate Φ_{I} with $\bar{\Phi}_{\mathrm{I}}$ and $\bar{\Phi}_{\mathrm{I}}$ is same as the parent process (which is a PPP) with co-located N daughters. The last equality is obtained from the probability generating functional (PGFL) of PPP. Similarly, using the **Model 2**, where we approximate Φ_{I} with $\bar{\bar{\Phi}}_{\mathrm{I}}$, **Part A** can be derived as follows:

$$\mathbb{E}_{\Phi_{\mathrm{I}}} \left[\prod_{x \in \Phi_{\mathrm{I}}} \left(\frac{1}{1 + \theta\|x_{(m)}\|^{\alpha}\|x\|^{-\alpha}} \right)^{b} \right] \approx \mathbb{E} \left[\prod_{x \in \bar{\bar{\Phi}}_{\mathrm{I}}} \left(\frac{1}{1 + \theta\|x_{(m)}\|^{\alpha}\|x\|^{-\alpha}} \right)^{b} \mid x_{(m)} \right]$$

$$= \exp \left(-2\pi N\lambda_{\mathrm{b}} \int_{0}^{\infty} \left[1 - \left(\frac{1}{1 + \theta\|x_{(m)}\|^{\alpha}r^{-\alpha}} \right)^{b} \right] \left(1 - e^{-(12/5)\lambda_{\mathrm{b}}\pi r^{2}} \right) r\mathrm{d}r \right).$$

Now, **Part B** in (A.2) is derived as follows:

$$\mathbb{E} \left[\left(\frac{1}{1 + \theta\|x_{(m)}\|^{\alpha}r^{-\alpha}} \right)^{b} \mid x_{(m)} \right] = \int_{\|x_{(m)}\|}^{\infty} \left(\frac{1}{1 + \theta\|x_{(m)}\|^{\alpha}r^{-\alpha}} \right)^{b} f_{R_{\mathrm{out}}} \left(r \mid \|x_{(m)}\| \right) \mathrm{d}r$$

$$= (5/2)\lambda_{\mathrm{b}}\pi\|x_{(m)}\|^{2} e^{(5/4)\lambda_{\mathrm{b}}\pi\|x_{(m)}\|^{2}} \int_{0}^{1} \frac{t^{-3}e^{-(5/4)\lambda_{\mathrm{b}}\pi\|x_{(m)}\|^{2}t^{-2}}}{(1 + \theta t^{\alpha})^{b}} \mathrm{d}t.$$

Finally, **Theorem 3.1** is obtained by averaging over the desired link distance using (3.8).

A.2 Proof of Theorem 3.2

Substituting (3.6) in (2.1), the CSP for the m-th rank user yields

$$P_{\mathrm{s},(m)}(\theta) = \mathbb{P} \left(\frac{I_{(m)}^{\mathrm{intra}} + I_{(m)}^{\mathrm{inter}}}{\beta_{m}P_{\mathrm{BS}}h_{0}\|x_{0}\|^{-\alpha}} < \frac{1}{\theta} \mid \Phi_{\mathrm{B}}, \mathrm{tx} \right)$$

$$\overset{(a)}{=} \mathbb{P} \left(\frac{I_{(m)}^{\mathrm{inter}}}{\beta_{m}P_{\mathrm{BS}}h_{0}\|x_{0}\|^{-\alpha}} < \frac{1}{\theta} - \frac{\sum_{i=1}^{m-1}\beta_{i}}{\beta_{m}} \mid \Phi_{\mathrm{B}}, \mathrm{tx} \right)$$

$$\overset{(b)}{=} \mathbb{E}_{h_{x}} \left[\exp \left\{ -c_{m}\|x_{0}\|^{\alpha} \left(\sum_{x \in \Phi_{\mathrm{B}} \setminus \{x_{0}\}} h_{x}\|x\|^{-\alpha} \right) \right\} \right]$$

119

$$= \prod_{x \in \Phi_{\mathrm{B}} \backslash \{x_0\}} \frac{1}{1 + c_m \|x_0\|^\alpha \|x\|^{-\alpha}},$$

where (a) is obtained by using (3.4). When $\beta_m/\theta - \sum_{i=1}^{m-1} \beta_i$ is not positive, $P_{\mathrm{s},(m)}(\theta) = 0$. Therefore, in the following, we consider $\beta_m/\theta - \sum_{i=1}^{m-1} \beta_i > 0$. (b) follows from the exponential distribution of h_0, applying (3.5), and setting $c_m = \left(\frac{\beta_m}{\theta} - \sum_{i-1}^{m-1} \beta_i \right)^{-1}$. Using $P_{\mathrm{s},(m)}$, now we can derive $M_{b,(m)}$ as follows:

$$M_{b,(m)} \overset{(a)}{=} \mathbb{E} \left[\prod_{x \in \Phi_{\mathrm{B}} \backslash \{x_0\}} \left(\frac{1}{1 + c_m \|x_0\|^\alpha \|x\|^{-\alpha}} \right)^b \right]$$

$$\overset{(b)}{=} \mathbb{E}_{R_m} \left[\exp \left\{ - \int_{\mathbb{R}^2 \backslash b(o,R_m)} \left[1 - \left(\frac{1}{1 + c_m R_m^\alpha \|x\|^{-\alpha}} \right)^b \right] \lambda_{\mathrm{b}} \mathrm{d}x \right\} \right]$$

$$\overset{(c)}{=} \mathbb{E}_{R_m} \left[\exp \left\{ -2\pi\lambda_{\mathrm{b}} \int_{R_m}^\infty \sum_{k=1}^\infty \binom{b}{k} (-1)^{k+1} c_m^k R_m^{\alpha k} \frac{r^{-\alpha k+1}}{(1 + c_m R_m^\alpha r^{-\alpha})^k} \mathrm{d}r \right\} \right]$$

$$\overset{(d)}{=} \int_0^\infty \exp \left\{ -\pi\lambda_{\mathrm{b}} r^2 \sum_{k=1}^\infty \binom{b}{k} (-1)^{k+1} c_m^k \frac{\delta}{k-\delta} \, {}_2F_1(k, k-\delta; k-\delta+1; -c_m) \right\} f_{R_m}(r) \mathrm{d}r,$$

where the expectation in (a) is over the point process Φ_{B}, (b) follows from probability generating functional (PGFL) of PPP [41] outside $b(o, R_m)$, (c) is obtained by using the polar domain representation and by applying the binomial expansion, and finally, $M_{b,(m)}$ in (3.21) is obtained by calculating the integral in (d) where $f_{R_m}(r)$ is given in (3.20).

Appendix B

Appendix to Chapter 4

B.1 Proof of (4.6)

In two user uplink NOMA, transmission success probability of the near user to the BS ($P_{\text{cov},(1)}^{\text{ISP}}$) is obtained by

$$
\begin{aligned}
P_{\text{cov},(1)}^{\text{ISP}} = {}& \mathbb{P}\left(\text{success of near user} \mid \text{received signal from near user is stronger than far user}\right) \\
& \times \mathbb{P}\left(\text{received signal from near user is stronger than far user}\right) \\
& + \mathbb{P}\left(\text{success of near user} \mid \text{received signal from near user is weaker than far user}\right) \\
& \times \mathbb{P}\left(\text{received signal from near user is weaker than far user}\right).
\end{aligned} \tag{B.1}
$$

We can write the second and fourth terms in the above equation as

$$
\mathbb{P}\left(\text{received signal from near user is stronger than far user}\right) = \mathbb{P}\left(h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha}\right)
$$

$$
\mathbb{P}\left(\text{received signal from near user is weaker than far user}\right) = \mathbb{P}\left(h_1 r_{(1)}^{-\alpha} < h_2 r_{(2)}^{-\alpha}\right)
$$

It is worth mentioning that, according to **Definition 4.1** in the thesis, second term in (B.1) is the (ranking) accuracy probability and fourth term is its complement.

Next, we simplify the first and third terms. Note that in the SIC process, receiver (BS) first decodes the strongest signal by treating other signals as noise. Then, it regenerates the decoded signal and subtracts the regenerated signal from the composite received signal and proceeds the SIC process on the remaining signal by decoding, regenerating, and canceling the second strongest signal and so on. Thus, when received power

of near user is stronger than far user, BS decodes near user's signal in the presence of interference from the far user, and we can write the first term as

$$\mathbb{P}\left(\text{success of near user} \mid \text{received signal from near user is stronger than far user}\right)$$

$$= \mathbb{P}\left\{ \frac{P_{\text{tx}}h_1 r_{(1)}^{-\alpha}}{P_{\text{tx}}h_2 r_{(2)}^{-\alpha} + I_{\text{inter}} + \sigma_n^2} > \theta \mid h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} \right\},$$

where, in the denominator of the above equation, the first term is due to the interference from the far (weaker) user, second term is due to the interference coming from transmissions in other cells (inter-cell interference), and the last term is the noise power.

On the other hand, when the near user is weaker than the far user at the BS, BS first decodes the far user's signal in the presence of interference from near user. Then it decodes the near user's signal (without interference from far user since signal of far user is removed in the SIC). Thus, we can write the third term as

$$\mathbb{P}\left(\text{success of near user} \mid \text{received signal from near user is weaker than far user}\right)$$

$$= \mathbb{P}\left\{ \frac{P_{\text{tx}}h_2 r_{(2)}^{-\alpha}}{P_{\text{tx}}h_1 r_{(1)}^{-\alpha} + I_{\text{inter}} + \sigma_n^2} > \theta, \frac{P_{\text{tx}}h_1 r_{(1)}^{-\alpha}}{I_{\text{inter}} + \sigma_n^2} > \theta \mid h_1 r_{(1)}^{-\alpha} < h_2 r_{(2)}^{-\alpha} \right\}.$$

Similarly, one can derive (4.7), (4.8), and (4.9).

B.2   Coverage Probability in Uplink NOMA With ISP-Based User Ranking

We consider 2-user uplink NOMA, and assume the network is interference-limited. Without loss of generality we can set $P_{\text{tx}} = 1$ in (4.6) and (4.7). Thus, we can write (4.6) and (4.7) as

$$P_{\text{cov},(1)}^{\text{ISP}} = \mathbb{P}\left\{ \frac{h_1 r_{(1)}^{-\alpha}}{h_2 r_{(2)}^{-\alpha} + I_{\text{inter}}} > \theta, h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} \right\} + \mathbb{P}\left\{ \frac{h_2 r_{(2)}^{-\alpha}}{h_1 r_{(1)}^{-\alpha} + I_{\text{inter}}} > \theta, \frac{h_1 r_{(1)}^{-\alpha}}{I_{\text{inter}}} > \theta, h_1 r_{(1)}^{-\alpha} < h_2 r_{(2)}^{-\alpha} \right\},$$

$$\tag{B.2}$$

$$P_{\text{cov},(2)}^{\text{ISP}} = \mathbb{P}\left\{ \frac{h_1 r_{(1)}^{-\alpha}}{h_2 r_{(2)}^{-\alpha} + I_{\text{inter}}} > \theta, \frac{h_2 r_{(2)}^{-\alpha}}{I_{\text{inter}}} > \theta, h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} \right\} + \mathbb{P}\left\{ \frac{h_2 r_{(2)}^{-\alpha}}{h_1 r_{(1)}^{-\alpha} + I_{\text{inter}}} > \theta, h_1 r_{(1)}^{-\alpha} < h_2 r_{(2)}^{-\alpha} \right\},$$

$$\tag{B.3}$$

where $I_{\text{inter}} = \sum_{x \in \Phi_I} h_x \|x\|^{-\alpha}$ is the inter-cell interference. According to Section 3.3, we can model Φ_I by a Poisson cluster process where the parent point process is a PPP with intensity function $\lambda \left(1 - e^{-(12/5)\lambda\pi\|x\|^2}\right)$, and, in each cluster, two offspring points are located in the same location as the parent.

As explained in Section 4.4, due to the complicated forms of (B.2) and (B.3), in the existing literature, it is assumed that the near user is always the strong user, i.e. $\mathbb{P}\left(h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha}\right) \approx 1$; therefore, (B.2) and (B.3) are approximated, respectively, by

$$P_{\text{cov},(1)}^{\text{MSP}} = \mathbb{P}\left\{\frac{h_1 r_{(1)}^{-\alpha}}{h_2 r_{(2)}^{-\alpha} + I_{\text{inter}}} > \theta\right\}, \tag{B.4}$$

$$P_{\text{cov},(2)}^{\text{MSP}} = \mathbb{P}\left\{\frac{h_1 r_{(1)}^{-\alpha}}{h_2 r_{(2)}^{-\alpha} + I_{\text{inter}}} > \theta, \frac{h_2 r_{(2)}^{-\alpha}}{I_{\text{inter}}} > \theta\right\}. \tag{B.5}$$

In the following, for Rayleigh fading, we calculate the above coverage probabilities and compare (B.2) and (B.3) with (B.4) and (B.5), respectively.

Theorem B.1. *(SIR distribution for ISP-based ranking) For Rayleigh fading, $P_{\text{cov},(1)}^{\text{ISP}}$ and $P_{\text{cov},(2)}^{\text{ISP}}$ can be obtained by*

$$P_{\text{cov},(1)}^{\text{ISP}} = \int_0^\infty \int_0^\infty P_{\text{cov},(1)|r_1,r_2}^{\text{ISP}} f_{r_{(1)},r_{(2)}}(r_1,r_2)\mathrm{d}r_1\mathrm{d}r_2$$

$$= \int_0^\infty \int_0^\infty \left(\frac{1}{1 + \theta\left(\frac{r_1}{r_2}\right)^\alpha}\mathcal{L}_{I_{\text{inter}}|r_1,r_2}\left(\theta r_1^\alpha\right) + \frac{1}{1 + \theta\left(\frac{r_2}{r_1}\right)^\alpha}\mathcal{L}_{I_{\text{inter}}|r_1,r_2}\left(\theta r_2^\alpha + \theta r_1^\alpha + \theta^2 r_2^\alpha\right)\right.$$

$$\left. + \mathbf{1}(\theta < 1)\left(1 - \frac{1}{1 + \theta\left(\frac{r_1}{r_2}\right)^\alpha} - \frac{1}{1 + \theta\left(\frac{r_2}{r_1}\right)^\alpha}\right)\mathcal{L}_{I_{\text{inter}}|r_1,r_2}\left(\frac{\theta}{1-\theta}r_1^\alpha + \frac{\theta}{1-\theta}r_2^\alpha\right)\right) f_{r_{(1)},r_{(2)}}(r_1,r_2)\mathrm{d}r_1\mathrm{d}r_2, \tag{B.6}$$

$$P_{\text{cov},(2)}^{\text{ISP}} = \int_0^\infty \int_0^\infty P_{\text{cov},(2)|r_1,r_2}^{\text{ISP}} f_{r_{(1)},r_{(2)}}(r_1,r_2)\mathrm{d}r_1\mathrm{d}r_2$$

$$= \int_0^\infty \int_0^\infty \left(\frac{1}{1 + \theta\left(\frac{r_2}{r_1}\right)^\alpha}\mathcal{L}_{I_{\text{inter}}|r_1,r_2}\left(\theta r_2^\alpha\right) + \frac{1}{1 + \theta\left(\frac{r_1}{r_2}\right)^\alpha}\mathcal{L}_{I_{\text{inter}}|r_1,r_2}\left(\theta r_1^\alpha + \theta r_2^\alpha + \theta^2 r_1^\alpha\right)\right.$$

$$\left. + \mathbf{1}(\theta < 1)\left(1 - \frac{1}{1 + \theta\left(\frac{r_1}{r_2}\right)^\alpha} - \frac{1}{1 + \theta\left(\frac{r_2}{r_1}\right)^\alpha}\right)\mathcal{L}_{I_{\text{inter}}|r_1,r_2}\left(\frac{\theta}{1-\theta}r_1^\alpha + \frac{\theta}{1-\theta}r_2^\alpha\right)\right) f_{r_{(1)},r_{(2)}}(r_1,r_2)\mathrm{d}r_1\mathrm{d}r_2, \tag{B.7}$$

where $P_{\text{cov},(1)|r_1,r_2}^{\text{ISP}}$ and $P_{\text{cov},(2)|r_1,r_2}^{\text{ISP}}$ denote the conditional coverage probabilities of near and far users given their distances from their serving BS. $f_{r_{(1)},r_{(2)}}(r_1,r_2)$ denotes the joint PDF of $r_{(1)}$ and $r_{(2)}$ and is given

in Definition 4.2. $\mathcal{L}_{I_{\text{Inter}}|r_1,r_2}(s) = \mathbb{E}\left[e^{-sI_{\text{inter}}} \mid r_{(1)} = r_1, r_{(2)} = r_2\right]$ also denotes the Laplace transform of the inter-cell interference given distances of the typical near and far users from their serving BS[1] From the PGFL of PPP, we have

$$\mathcal{L}_{I_{\text{Inter}}|r_1,r_2}(s) = \exp\left\{-2\pi\lambda \int_0^\infty \left(1 - (1 + sx^{-\alpha})^{-2}\right)\left(1 - e^{-(12/5)\lambda\pi x^2}\right)x\mathrm{d}x\right\}.$$

Proof. We only provide the proof for $P_{\text{cov},(1)}^{\text{ISP}}$; $P_{\text{cov},(2)}^{\text{ISP}}$ can be proven following the same steps.

The first term in (B.2) can be obtained by

$$\mathbb{P}\left\{\frac{h_1 r_{(1)}^{-\alpha}}{h_2 r_{(2)}^{-\alpha} + I_{\text{inter}}} > \theta, h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha}\right\} = \mathbb{P}\left\{h_1 > \max\left\{\theta r_{(1)}^\alpha\left(h_2 r_{(2)}^{-\alpha} + I_{\text{inter}}\right), h_2\left(\frac{r_{(1)}}{r_{(2)}}\right)^\alpha\right\}\right\}$$

$$= \mathbb{E}\left[\mathbf{1}\left(h_1 > \theta r_{(1)}^\alpha\left(h_2 r_{(2)}^{-\alpha} + I_{\text{inter}}\right)\right)\right]$$

$$- \mathbf{1}(\theta < 1)\mathbb{E}\left[\mathbf{1}\left(\frac{h_2 r_{(2)}^{-\alpha}}{I_{\text{inter}}} > \frac{\theta}{1-\theta}\right)\mathbf{1}\left(h_1 > \theta r_{(1)}^\alpha\left(h_2 r_{(2)}^{-\alpha} + I_{\text{inter}}\right)\right)\right]$$

$$+ \mathbf{1}(\theta < 1)\mathbb{E}\left[\mathbf{1}\left(\frac{h_2 r_{(2)}^{-\alpha}}{I_{\text{inter}}} > \frac{\theta}{1-\theta}\right)\mathbf{1}\left(h_1 > h_2\left(\frac{r_{(1)}}{r_{(2)}}\right)^\alpha\right)\right]. \tag{B.8}$$

For the second term in (B.2), we also have

$$\mathbb{P}\left\{\frac{h_2 r_{(2)}^{-\alpha}}{h_1 r_{(1)}^{-\alpha} + I_{\text{inter}}} > \theta, \frac{h_1 r_{(1)}^{-\alpha}}{I_{\text{inter}}} > \theta, h_1 r_{(1)}^{-\alpha} < h_2 r_{(2)}^{-\alpha}\right\}$$

$$= \mathbb{P}\left\{h_2 > \max\left\{\theta r_{(2)}^\alpha\left(h_1 r_{(1)}^{-\alpha} + I_{\text{inter}}\right), h_1\left(\frac{r_{(2)}}{r_{(1)}}\right)^\alpha\right\}, \frac{h_1 r_{(1)}^{-\alpha}}{I_{\text{inter}}} > \theta\right\}$$

$$= \mathbb{E}\left[\mathbf{1}\left(h_2 > \theta r_{(2)}^\alpha\left(h_1 r_{(1)}^{-\alpha} + I_{\text{inter}}\right)\right)\mathbf{1}\left(\frac{h_1 r_{(1)}^{-\alpha}}{I_{\text{inter}}} > \theta\right)\right]$$

$$- \mathbf{1}(\theta < 1)\mathbb{E}\left[\mathbf{1}\left(\frac{h_1 r_{(1)}^{-\alpha}}{I_{\text{inter}}} > \frac{\theta}{1-\theta}\right)\mathbf{1}\left(h_2 > \theta r_{(2)}^\alpha\left(h_1 r_{(1)}^{-\alpha} + I_{\text{inter}}\right)\right)\right]$$

$$+ \mathbf{1}(\theta < 1)\mathbb{E}\left[\mathbf{1}\left(\frac{h_1 r_{(1)}^{-\alpha}}{I_{\text{inter}}} > \frac{\theta}{1-\theta}\right)\mathbf{1}\left(h_2 > h_1\left(\frac{r_{(2)}}{r_{(1)}}\right)^\alpha\right)\right]. \tag{B.9}$$

Finally, $P_{\text{cov},(1)}^{\text{ISP}}$ can be obtained by taking the expectation in (B.8) and (B.9) with respect to h_1 and h_2. $\square$

Following the same steps as in proof of **Theorem B.1**, for $P_{\text{cov},(1)}^{\text{MSP}}$ and $P_{\text{cov},(2)}^{\text{MSP}}$, we obtain

$$rClP_{\text{cov},(1)}^{\text{MSP}} = \int_0^\infty \int_0^\infty \frac{1}{1 + \theta\left(\frac{r_1}{r_2}\right)^\alpha}\mathcal{L}_{I_{\text{inter}}|r_1,r_2}(\theta r_1^\alpha)f_{r_{(1)},r_{(2)}}(r_1, r_2)\mathrm{d}r_1\mathrm{d}r_2, \tag{B.10}$$

---

[1]Note that in our proposed models, I_{inter} is independent of $r_{(1)}$ and $r_{(2)}$.

$$P_{\text{cov},(2)}^{\text{MSP}} = \int_0^\infty \int_0^\infty \frac{1}{1 + \theta \left(\frac{r_1}{r_2}\right)^\alpha} \mathcal{L}_{I_{\text{inter}}|r_1,r_2}(\theta r_1^\alpha + \theta r_2^\alpha + \theta^2 r_1^\alpha) f_{r_{(1)},r_{(2)}}(r_1,r_2) \mathrm{d}r_1 \mathrm{d}r_2. \tag{B.11}$$

By comparing (B.6) with (B.10) and (B.7) with (B.11), we see that, the derived results in the existing literature, which are obtained using MSP-based user ranking underestimate the coverage probability. This is proved in the following corollary.

Corollary B.1. *For Rayleigh fading, the MSP-based user ranking provides a lower bound for the coverage probability, i.e. $P_{\text{cov},(i)}^{\text{MSP}} \leq P_{\text{cov},(i)}^{\text{ISP}}$, $i \in \{1,2\}$.*

Proof. We only consider $\theta < 1$, since for $\theta \geq 1$, the proof is straightforward.

Since $\mathcal{L}_{I_{\text{Inter}}|r_1,r_2}(s)$ is a decreasing function of s, we have

$$\mathcal{L}_{I_{\text{inter}}|r_1,r_2}\left(\frac{\theta}{1-\theta}r_1^\alpha + \frac{\theta}{1-\theta}r_2^\alpha\right) = \mathcal{L}_{I_{\text{inter}}|r_1,r_2}\left(\theta r_2^\alpha + \frac{\theta}{1-\theta}r_1^\alpha + \frac{\theta^2}{1-\theta}r_2^\alpha\right)$$

$$\leq \mathcal{L}_{I_{\text{inter}}|r_1,r_2}\left(\theta r_2^\alpha + \theta r_1^\alpha + \theta^2 r_2^\alpha\right).$$

Using the above inequality besides $\frac{1}{1+\theta r_1^\alpha r_2^{-\alpha}} \leq 1$, we get $P_{\text{cov},(1)}^{\text{MSP}} \leq P_{\text{cov},(1)}^{\text{ISP}}$. A similar approach can be used for the far user. $\qquad\square$

B.3   Proof of Theorem 4.1

For Rayleigh fading and N-UE NOMA, the inner expectation over $\{h_i\}_{i=1}^N$ can be derived as:

$$\mathbb{E}_{\{h_i\}}\left[\mathbf{1}\left(h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} > \cdots > h_N r_{(N)}^{-\alpha}\right)\right]$$

$$= \mathbb{E}\left[\mathbf{1}\left(h_2 r_{(2)}^{-\alpha} > \cdots > h_N r_{(N)}^{-\alpha}\right)\mathbb{E}_{h_1}\left[\mathbf{1}\left(h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha}\right)\right]\right]$$

$$\stackrel{(a)}{=} \mathbb{E}\left[\mathbf{1}\left(h_2 r_{(2)}^{-\alpha} > \cdots > h_N r_{(N)}^{-\alpha}\right)\exp\left\{-\frac{h_2\left(\frac{r_{(1)}}{r_{(2)}}\right)^\alpha}{\Omega}\right\}\right]$$

$$= \mathbb{E}\left[\mathbf{1}\left(h_3 r_{(3)}^{-\alpha} > \cdots > h_N r_{(N)}^{-\alpha}\right)\mathbb{E}_{h_2}\left[\mathbf{1}\left(h_2 r_{(2)}^{-\alpha} > h_3 r_{(3)}^{-\alpha}\right)\exp\left\{-\frac{h_2\left(\frac{r_{(1)}}{r_{(2)}}\right)^\alpha}{\Omega}\right\}\right]\right]$$

$$\stackrel{(b)}{=} \mathbb{E}\left[\mathbf{1}\left(h_3 r_{(3)}^{-\alpha} > \cdots > h_N r_{(N)}^{-\alpha}\right)\frac{1}{1+\left(\frac{r_{(1)}}{r_{(2)}}\right)^\alpha}\exp\left\{-\frac{h_3\left[\left(\frac{r_{(1)}}{r_{(3)}}\right)^\alpha + \left(\frac{r_{(2)}}{r_{(3)}}\right)^\alpha\right]}{\Omega}\right\}\right]$$

$$\stackrel{(c)}{=} \frac{1}{1+\left(\frac{r_{(1)}}{r_{(2)}}\right)^\alpha}\cdot\frac{1}{1+\left(\frac{r_{(1)}}{r_{(3)}}\right)^\alpha + \left(\frac{r_{(2)}}{r_{(3)}}\right)^\alpha}\cdots\cdots\frac{1}{1+\left(\frac{r_{(1)}}{r_{(N)}}\right)^\alpha + \left(\frac{r_{(2)}}{r_{(N)}}\right)^\alpha\cdots\left(\frac{r_{(N-1)}}{r_{(N)}}\right)^\alpha}$$

$$= \prod_{i=2}^{N} \frac{1}{\sum_{j=1}^{i} \left(\frac{r_{(j)}}{r_{(i)}} \right)^{\alpha}},$$

where (a), (b), and (c) follow since $\{h_i\}$ are i.i.d. exponential random variables with mean Ω.

B.4   Proof of Theorem 4.2

Using PDF of the link distance (4.1) in (4.13) yields

$$\mathbb{E}\left[\left(\frac{r_{(1)}}{r_{(2)}} \right)^{\alpha k} \right] = 2 \int_0^{\infty} \int_{r_1}^{\infty} \left(\frac{r_1}{r_2} \right)^{\alpha k} f_r(r_1) f_r(r_2) \mathrm{d}r_2 \mathrm{d}r_1$$

$$= 2(2c\lambda\pi)^2 \int_0^{\infty} \int_{r_1}^{\infty} \left(\frac{r_1}{r_2} \right)^{\alpha k} r_1 r_2 e^{-c\lambda\pi(r_1^2+r_2^2)} \mathrm{d}r_2 \mathrm{d}r_1.$$

Applying changes of variables $\frac{r_1}{r_2} = u$ and $r_1 = v$, we have

$$\mathbb{E}\left[\left(\frac{r_{(1)}}{r_{(2)}} \right)^{\alpha k} \right] = 2(2c\lambda\pi)^2 \int_0^{\infty} \int_0^1 u^{\alpha k-3} v^3 e^{-c\lambda\pi(1+\frac{1}{u^2})v^2} \mathrm{d}u \mathrm{d}v$$

$$= 2(2c\lambda\pi)^2 \int_0^1 u^{\alpha k-3} \int_0^{\infty} v^3 e^{-c\lambda\pi(1+\frac{1}{u^2})v^2} \mathrm{d}v \mathrm{d}u.$$

Applying $c\lambda\pi(1+\frac{1}{u^2})v^2 = t$ in the inner integral yields

$$\mathbb{E}\left[\left(\frac{r_{(1)}}{r_{(2)}} \right)^{\alpha k} \right] = 2(2c\lambda\pi)^2 \int_0^1 u^{\alpha k-3} \int_0^{\infty} \frac{te^{-t}\mathrm{d}t}{2(c\lambda\pi)^2(1+\frac{1}{u^2})^2} \mathrm{d}u = 4\int_0^1 \frac{u^{\alpha k+1}\mathrm{d}u}{(1+u^2)^2} = 2\int_0^1 \frac{x^{\alpha k/2}\mathrm{d}x}{(1+x)^2}$$

$$= \frac{2}{\frac{\alpha k}{2}+1} \,_2F_1\left(2, \frac{\alpha k}{2}+1; \frac{\alpha k}{2}+2; -1 \right).$$

Finally, Theorem 4.2 is obtained by $_2F_1(a,b;c;z) = (1-z)^{-a} \,_2F_1(a,c-b;c;\frac{z}{z-1})$.

B.5   Proof of Corollary 4.3

From substituting (4.1) in (4.11), we get

$$\mathcal{A} = N!(2c\lambda\pi)^N \int_0^{\infty} \int_{r_1}^{\infty} \cdots \int_{r_{N-1}}^{\infty} r_1 e^{-c\lambda\pi r_1^2} \prod_{i=2}^{N} \frac{r_i e^{-c\lambda\pi r_i^2}}{\sum_{j=1}^{i} \left(\frac{r_j}{r_i} \right)^{\alpha}} \mathrm{d}r_N \cdots \mathrm{d}r_2 \mathrm{d}r_1. \tag{B.12}$$

Next, we simplify (B.12) by applying changes of variables $r_1 = u_1$, $\frac{r_{i-1}}{r_i} = u_i$ for $i = 2, ..., N$. Since we have $r_i = \frac{u_1}{u_2 u_3 \cdots u_i}$, $i = 2, ..., N$, the Jacobian matrix $J = \frac{\partial(r_1, \cdots, r_N)}{\partial(u_1, \cdots, u_N)}$ is a triangular matrix and its determinant is

equal to the product of the main diagonal entries, i.e., $\det(J) = \prod_{i=1}^{N} \frac{\partial r_i}{\partial u_i} = \prod_{i=2}^{N} \frac{-u_1}{u_2 u_3 \cdots u_{i-1} u_i^2}$. Moreover, after changes of variables the region of integration is as $0 < u_1$ and $0 < u_i < 1$ for $i = 2, ..., N$. Therefore, (B.12) can be written as follows:

$$
\mathcal{A} = N!(2c\lambda\pi)^N \int_0^\infty \int_0^1 \cdots \int_0^1 u_1 e^{-c\lambda\pi u_1^2} \prod_{i=2}^{N} \left(\frac{u_1^2 u_i^{-1} \prod_{k=2}^{i} u_k^{-2}}{1 + \sum_{j=1}^{i-1} \prod_{n=j+1}^{i} u_n^\alpha} e^{-c\lambda\pi \frac{u_1^2}{\prod_{m=2}^{i} u_m^2}} \right) du_N \cdots du_2 du_1
$$

$$
= N!(2c\lambda\pi)^N \int_0^1 \cdots \int_0^1 \prod_{i=2}^{N} \frac{u_i^{-1} \prod_{k=2}^{i} u_k^{-2}}{1 + \sum_{j=1}^{i-1} \prod_{n=j+1}^{i} u_n^\alpha} \int_0^\infty u_1^{2N-1} e^{-c\lambda\pi \left(1 + \sum_{i=2}^{N} \prod_{m=2}^{i} u_m^{-2}\right) u_1^2} du_1 \, du_N \cdots du_2.
$$

Finally, rewriting $\prod_{i=2}^{N} u_i^{-1} \prod_{k=2}^{i} u_k^{-2}$ as $\prod_{i=2}^{N} u_i^{-3-2(N-i)}$ and applying $c\lambda\pi \left(1 + \sum_{i=2}^{N} \prod_{m=2}^{i} u_m^{-2}\right) u_1^2 = t$ yields

$$
\mathcal{A} = N!(2c\lambda\pi)^N \int_0^1 \cdots \int_0^1 \prod_{i=2}^{N} \frac{u_i^{-3-2(N-i)}}{1 + \sum_{j=1}^{i-1} \prod_{n=j+1}^{i} u_n^\alpha} \int_0^\infty \frac{t^{N-1} e^{-t} dt}{2 \left(c\lambda\pi \left[1 + \sum_{i=2}^{N} \prod_{m=2}^{i} u_m^{-2}\right]\right)^N} du_N \cdots du_2
$$

$$
= N!(N-1)! \, 2^{N-1} \int_0^1 \cdots \int_0^1 \frac{1}{\left(1 + \sum_{i=2}^{N} \prod_{m=2}^{i} u_m^{-2}\right)^N} \prod_{i=2}^{N} \frac{u_i^{-3-2(N-i)}}{1 + \sum_{j=1}^{i-1} \prod_{n=j+1}^{i} u_n^\alpha} du_N \cdots du_2, \quad (B.13)
$$

where the final equation is obtained by using the definition of the gamma function. According to (B.13), $\mathcal{A}$ depends on α and N; it does not depend on the BS intensity λ.

B.6 Proof of Theorem 4.3

For Nakagami-m fading and N-UE NOMA, the inner expectation over $\{h\}_{i=1}^{N}$ can be derived as:

$$
\mathbb{E}_{\{h_i\}} \left[\mathbf{1} \left(h_1 r_{(1)}^{-\alpha} > h_2 r_{(2)}^{-\alpha} > \cdots > h_N r_{(N)}^{-\alpha} \right) \right]
$$

$$
= \mathbb{E} \left[\mathbf{1} \left(h_2 r_{(2)}^{-\alpha} > \cdots > h_N r_{(N)}^{-\alpha} \right) \int_{h_2 \left(\frac{r_{(1)}}{r_{(2)}} \right)^\alpha}^\infty \frac{m^m h_1^{m-1}}{\Gamma(m)\Omega^m} e^{-\frac{m}{\Omega} h_1} dh_1 \right]
$$

$$
\overset{(a)}{=} \mathbb{E} \left[\mathbf{1} \left(h_2 r_{(2)}^{-\alpha} > \cdots > h_N r_{(N)}^{-\alpha} \right) \frac{m^m}{\Gamma(m)\Omega^m} h_2^m \left(\frac{r_{(1)}}{r_{(2)}} \right)^{\alpha m} \int_1^\infty t_1^{m-1} e^{-\frac{m}{\Omega} t_1 h_2 \left(\frac{r_{(1)}}{r_{(2)}} \right)^\alpha} dt_1 \right]
$$

$$
= \mathbb{E} \left[\mathbf{1} \left(h_3 r_{(3)}^{-\alpha} > \cdots > h_N r_{(N)}^{-\alpha} \right) \left(\frac{m^m}{\Gamma(m)\Omega^m} \right)^2 \left(\frac{r_{(1)}}{r_{(2)}} \right)^{\alpha m} \right.
$$

$$
\left. \cdot \int_1^\infty t_1^{m-1} \int_{h_3 \left(\frac{r_{(2)}}{r_{(3)}} \right)^\alpha}^\infty h_2^{2m-1} \exp \left\{ -\frac{m}{\Omega} h_2 \left[1 + t_1 \left(\frac{r_{(1)}}{r_{(2)}} \right)^\alpha \right] \right\} dh_2 dt_1 \right]
$$

$$
\overset{(b)}{=} \mathbb{E} \left[\mathbf{1} \left(h_3 r_{(3)}^{-\alpha} > \cdots > h_N r_{(N)}^{-\alpha} \right) \left(\frac{m^m}{\Gamma(m)\Omega^m} \right)^2 \left(\frac{r_{(1)}}{r_{(2)}} \right)^{\alpha m} \left(\frac{r_{(2)}}{r_{(3)}} \right)^{2\alpha m} h_3^{2m} \right.
$$

$$\cdot \int_1^\infty t_1^{m-1} \int_1^\infty t_2^{2m-1} \exp\left\{-\frac{m}{\Omega}h_3\left[t_2\left(\frac{r_{(2)}}{r_{(3)}}\right)^\alpha + t_1 t_2\left(\frac{r_{(1)}}{r_{(3)}}\right)^\alpha\right]\right\}dt_2 dt_1\Bigg]$$

$$\overset{(c)}{=} \left(\frac{m^m}{\Gamma(m)\Omega^m}\right)^N \left(\frac{r_{(1)}}{r_{(2)}}\right)^{\alpha m}\left(\frac{r_{(2)}}{r_{(3)}}\right)^{2\alpha m}\cdots\left(\frac{r_{(N-1)}}{r_{(N)}}\right)^{(N-1)\alpha m} \int_1^\infty \int_1^\infty \cdots \int_1^\infty t_1^{m-1}t_2^{2m-1}\cdots t_{N-1}^{(N-1)m-1}$$

$$\cdot \int_0^\infty h_N^{Nm-1}\exp\left\{-\frac{m}{\Omega}h_N\left[1+\sum_{i=1}^{N-1}\left(\frac{r_{(i)}}{r_{(N)}}\right)^\alpha \prod_{k=i}^{N-1}t_k\right]\right\}dh_N dt_{N-1}\cdots dt_2 dt_1.$$

where (a), (b) are obtained by changes of variables $h_1 = h_2\left(\frac{r_{(1)}}{r_{(2)}}\right)^\alpha t_1$, and $h_2 = h_3\left(\frac{r_{(2)}}{r_{(3)}}\right)^\alpha t_2$. (c) follows by averaging over $h_3, ..., h_N$. Finally, **Theorem 4.3** can be obtained by applying $t_N = \frac{m}{\Omega}h_N\left[1+\sum_{i=1}^{N-1}\left(\frac{r_{(i)}}{r_{(N)}}\right)^\alpha \prod_{k=i}^{N-1}t_k\right]$.

B.7   Proof of Corollary 4.4

From (4.4) and **Theorem 4.3**, we can derive $\mathcal{A}$ as follows:

$$\mathcal{A} = \frac{\Gamma(Nm)}{\Gamma(m)^N}\int_1^\infty \int_1^\infty \cdots \int_1^\infty \mathbb{E}\left[\frac{\prod_{j=1}^{N-1}\left(\frac{r_{(j)}}{r_{(N)}}\right)^{\alpha m}}{\left[1+\sum_{i=1}^{N-1}\left(\frac{r_{(i)}}{r_{(N)}}\right)^\alpha \prod_{k=i}^{N-1}t_k\right]^{Nm}}\right]\prod_{j=1}^{N-1}t_j^{jm-1}dt_{N-1}\cdots dt_2 dt_1, \quad \text{(B.14)}$$

where expectation is over $\{r_{(i)}\}$. Following the same steps as **Corollary 4.3**, we have

$$\mathbb{E}\left[\frac{\prod_{j=1}^{N-1}\left(\frac{r_{(j)}}{r_{(N)}}\right)^{\alpha m}}{\left[1+\sum_{i=1}^{N-1}\left(\frac{r_{(i)}}{r_{(N)}}\right)^\alpha \prod_{k=i}^{N-1}t_k\right]^{Nm}}\right] =$$

$$N!(N-1)!\,2^{N-1}\int_0^1 \cdots \int_0^1 \frac{\prod_{j=2}^{N}u_j^{-3-2(N-j)+(j-1)\alpha m}}{\left(1+\sum_{i=1}^{N-1}\prod_{k=i}^{N-1}u_{k+1}^\alpha t_k\right)^{Nm}\left(1+\sum_{i=2}^{N}\prod_{k=2}^{i}u_k^{-2}\right)^N}du_N\cdots du_2,$$

which is independent of λ.

B.8   Proof of Corollary 4.5

Assume that M users are associated to the typical BS. Here we have used $[M]$ to denote the set $\{1, 2, \cdots, M\}$, and we have used the notation $\mathbb{P}\left(h_{\mathcal{U}_{\min}}r_{(\mathcal{U}_{\min})}^{-\alpha} > h_{\mathcal{U}_{\max}}r_{(\mathcal{U}_{\max})}^{-\alpha}\mid \mathcal{U} = [M]\right)$ to emphasize that the NOMA cluster is formed by selecting the nearest and the farthest users from the set $\mathcal{U}$, where $\mathcal{U}$ includes ranks of users that we are allowed to select, $\mathcal{U}_{\min} = \min \mathcal{U}$ is the rank of the nearest user, and $\mathcal{U}_{\max} = \max \mathcal{U}$ is the rank of the farthest user. Note that removing one user from the set $[M]$ corresponds to the case that $M-1$ users are associated to the typical BS. Let assume that user i is removed from the set $[M]$, and the NOMA cluster is formed by selecting the nearest and the farthest user from the set $[M] \setminus \{i\}$. Based on value of i,

two different scenarios can occur:

1) When $i \in \{2, \cdots, M-1\}$, the nearest and the farthest users in the set $[M] \setminus \{i\}$ are users at rank 1 and M, respectively; therefore,

$$\mathbb{P}\left(h_{\mathcal{U}_{\min}} r_{(\mathcal{U}_{\min})}^{-\alpha} > h_{\mathcal{U}_{\max}} r_{(\mathcal{U}_{\max})}^{-\alpha} \,|\, \mathcal{U} = [M] \setminus \{i\}\right) = \mathbb{P}\left(h_{\mathcal{U}_{\min}} r_{(\mathcal{U}_{\min})}^{-\alpha} > h_{\mathcal{U}_{\max}} r_{(\mathcal{U}_{\max})}^{-\alpha} \,|\, \mathcal{U} = [M]\right).$$

2) When $i \in \{1, M\}$, we can show

$$\mathbb{P}\left(h_{\mathcal{U}_{\min}} r_{(\mathcal{U}_{\min})}^{-\alpha} > h_{\mathcal{U}_{\max}} r_{(\mathcal{U}_{\max})}^{-\alpha} \,|\, \mathcal{U} = [M] \setminus \{i\}\right) \leq \mathbb{P}\left(h_{\mathcal{U}_{\min}} r_{(\mathcal{U}_{\min})}^{-\alpha} > h_{\mathcal{U}_{\max}} r_{(\mathcal{U}_{\max})}^{-\alpha} \,|\, \mathcal{U} = [M]\right).$$

Therefore, when we select the nearest and the farthest user for the NOMA cluster, the accuracy probability increases as the size of set of users that we are selecting from increases.

To complete the proof we need to show $\mathbb{P}\left(h_i r_{(i)}^{-\alpha} > h_j r_{(j)}^{-\alpha} \,|\, [M]\right) \leq \mathbb{P}\left(h_1 r_{(1)}^{-\alpha} > h_M r_{(M)}^{-\alpha} \,|\, [M]\right)$, where $i, j \in [M]$. From (4.4), we have

$$\begin{aligned}
\mathbb{P}\left(h_i r_{(i)}^{-\alpha} > h_j r_{(j)}^{-\alpha} \,|\, [M]\right) &= \mathbb{E}_{h_i, h_j}\left[\mathbb{E}_{r_{(i)}, r_{(j)}}\left[\mathbf{1}\left(h_i r_{(i)}^{-\alpha} > h_j r_{(j)}^{-\alpha}\right)\right]\right] \\
&\overset{(a)}{\leq} \mathbb{E}_{h_i, h_j}\left[\mathbb{E}_{r_{(1)}, r_{(M)}}\left[\mathbf{1}\left(h_i r_{(1)}^{-\alpha} > h_j r_{(M)}^{-\alpha}\right)\right]\right] \\
&\overset{(b)}{=} \mathbb{E}_{h_1, h_M}\left[\mathbb{E}_{r_{(1)}, r_{(M)}}\left[\mathbf{1}\left(h_1 r_{(1)}^{-\alpha} > h_M r_{(M)}^{-\alpha}\right)\right]\right] \\
&= \mathbb{P}\left(h_1 r_{(1)}^{-\alpha} > h_M r_{(M)}^{-\alpha} \,|\, [M]\right),
\end{aligned}$$

where (a) is obtained since for any realization of user point process that $h_i r_{(i)}^{-\alpha} > h_j r_{(j)}^{-\alpha}$ is satisfied, $h_i r_{(1)}^{-\alpha} > h_j r_{(M)}^{-\alpha}$ is also true, i.e., $h_i r_{(i)}^{-\alpha} > h_j r_{(j)}^{-\alpha}$ is a sufficient condition for $h_i r_{(1)}^{-\alpha} > h_j r_{(M)}^{-\alpha}$. (b) follows since all the channel power gains are i.i.d.

Appendix C

Appendix to Chapter 5

C.1 Proof of Proposition 5.1

From (5.8), we have $\mathcal{B}\left(\mathrm{x}(0), \frac{r_0}{\beta_{kj}}\right) \cup \mathcal{B}\left(\mathrm{x}(T), \frac{r_0(T)}{\beta_{kj}}\right) \subset \mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj})$. To complete the proof we need to show that $\mathcal{A}_{kj}(r_0, \theta, v, T, \beta_{kj}) \subset \mathcal{B}\left(\mathrm{x}(0), \frac{r_0}{\beta_{kj}}\right) \cup \mathcal{B}\left(\mathrm{x}(T), \frac{r_0(T)}{\beta_{kj}}\right)$.

Consider a point $\mathrm{y} \in \mathcal{B}\left(\mathrm{x}(t), \frac{r_0(t)}{\beta_{kj}}\right)$, $0 \leq t \leq T$. Let us represent y in polar coordinates as (ζ, φ), where ζ is the distance between y and the origin $(\mathrm{x}(0))$ and φ is the angle made between the line segment from the origin to y and the positive x-axis (user's trajectory) (Fig. C.1). Using triangle equations, we have $\|\mathrm{y} - \mathrm{x}(t)\| = \sqrt{\zeta^2 + v^2 t^2 - 2\zeta v t \cos \varphi}$. Since $\mathrm{y} \in \mathcal{B}\left(\mathrm{x}(t), \frac{r_0(t)}{\beta_{kj}}\right)$, $0 \leq t \leq T$,

$$\zeta^2 + v^2 t^2 - 2\zeta v t \cos \varphi \leq \frac{r_0(t)^2}{\beta_{kj}^2} = \frac{r_0^2 + v^2 t^2 - 2r_0 v t \cos \theta}{\beta_{kj}^2}.$$

Rewriting the above inequality gives

$$\zeta^2 - \frac{r_0^2}{\beta_{kj}^2} \leq \left(\frac{1}{\beta_{kj}^2} - 1\right) v^2 t^2 + 2vt \left(\zeta \cos \varphi - \frac{r_0 \cos \theta}{\beta_{kj}^2}\right), \qquad 0 \leq t \leq T.$$

For $\beta_{kj} < 1$ $(0 < \frac{1}{\beta_{kj}^2} - 1)$, the right hand side of the above inequality is a convex function with respect to t. When $x \in [a, b]$, for a convex function f, we have $f(x) \leq \max\{f(a), f(b)\}$. Using this property of convex functions yields,

$$\zeta^2 - \frac{r_0^2}{\beta_{kj}^2} \leq \max \left\{0, \left(\frac{1}{\beta_{kj}^2} - 1\right) v^2 T^2 + 2vT \left(\zeta \cos \varphi - \frac{r_0 \cos \theta}{\beta_{kj}^2}\right)\right\}.$$
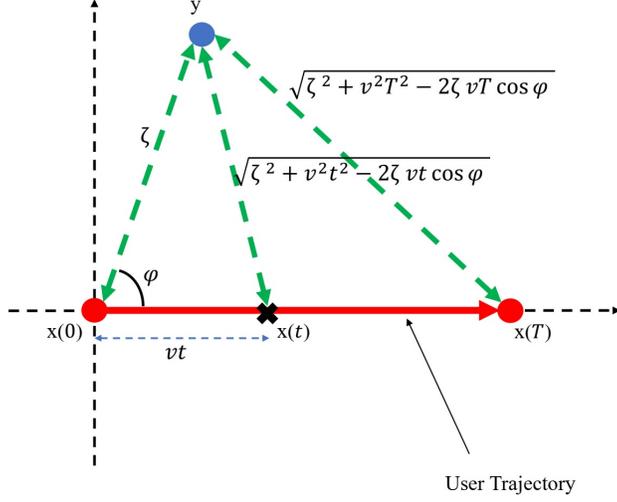
Figure C.1: A geometric illustration for proof of Proposition 1.

Therefore, depending on the parameters, we have

$$\zeta^2 - \frac{r_0^2}{\beta_{kj}^2} \leq 0, \quad \text{or} \quad \zeta^2 - \frac{r_0^2}{\beta_{kj}^2} \leq \left(\frac{1}{\beta_{kj}^2} - 1\right) v^2 T^2 + 2vT \left(\zeta \cos\varphi - \frac{r_0 \cos\theta}{\beta_{kj}^2}\right). \tag{C.1}$$

We can rewrite the inequalities in (C.1) as

$$\zeta^2 \leq \frac{r_0^2}{\beta_{kj}^2}, \quad \text{or} \quad \zeta^2 + v^2 T^2 - 2\zeta vT \cos\varphi \leq \frac{r_0^2 + v^2 T^2 - 2r_0 vT \cos\theta}{\beta_{kj}^2}. \tag{C.2}$$

$\zeta^2 \leq \frac{r_0^2}{\beta_{kj}^2}$ is equivalent to $\|y - x(0)\| \leq \frac{r_0}{\beta_{kj}}$, and $\zeta^2 + v^2 T^2 - 2\zeta vT \cos\varphi \leq \frac{r_0^2 + v^2 T^2 - 2r_0 vT \cos\theta}{\beta_{kj}^2}$ is equivalent to $\|y - x(T)\| \leq \frac{r_0(T)}{\beta_{kj}}$. Thus, $y \in \mathcal{B}\left(x(0), \frac{r_0}{\beta_{kj}}\right) \cup \mathcal{B}\left(x(T), \frac{r_0(T)}{\beta_{kj}}\right)$.

C.2 Proof of (5.19)

Before proving (5.19), we provide the Taylor series expansion of arccos and arcsin. They help us to derive the final result in (5.19).

The arcsin function has a Taylor expansion:

$$\arcsin(x) = \sum_{n=0}^{\infty} \frac{(2n)!}{2^{2n}(n!)^2} \frac{x^{2n+1}}{2n+1}, \tag{C.3}$$

By taking derivative with respect to x from both sides of (C.3), we get

$$\frac{1}{\sqrt{1-x^2}} = \sum_{n=0}^{\infty} \frac{(2n)!}{2^{2n}(n!)^2} x^{2n}. \tag{C.4}$$

131

Using $\arccos(x) = \frac{\pi}{2} - \arcsin(x)$, we can write

$$\arccos(x) = \frac{\pi}{2} - \sum_{n=0}^{\infty} \frac{(2n)!}{2^{2n}(n!)^2} \frac{x^{2n+1}}{2n+1}. \tag{C.5}$$

When $\mathrm{d}x \to 0$,

$$
\begin{aligned}
\arccos(x + \mathrm{d}x) &= \frac{\pi}{2} - \sum_{n=0}^{\infty} \frac{(2n)!}{2^{2n}(n!)^2} \frac{(x+\mathrm{d}x)^{2n+1}}{2n+1} \\
&= \frac{\pi}{2} - \sum_{n=0}^{\infty} \frac{(2n)!}{2^{2n}(n!)^2} \frac{x^{2n+1} + (2n+1)x^{2n}\mathrm{d}x + O(\mathrm{d}x^2)}{2n+1} \\
&= \frac{\pi}{2} - \sum_{n=0}^{\infty} \frac{(2n)!}{2^{2n}(n!)^2} \frac{x^{2n+1}}{2n+1} - \sum_{n=0}^{\infty} \frac{(2n)!}{2^{2n}(n!)^2} x^{2n}\mathrm{d}x + O(\mathrm{d}x^2) \\
&\overset{(a)}{=} \arccos(x) - \frac{\mathrm{d}x}{\sqrt{1-x^2}} + O(\mathrm{d}x^2),
\end{aligned}
\tag{C.6}
$$

where (a) is obtained using (C.4) and (C.5).

From (C.6), we obtain

$$
\begin{aligned}
&\arccos\left(\frac{r_0 \cos\theta - vt}{\beta_{kj} r_0(t)} + \frac{\beta_{kj}^2 - 1}{2\beta_{kj}} \frac{v\mathrm{d}t}{r_0(t)} \right) = \\
&\arccos\left(\frac{r_0 \cos\theta - vt}{\beta_{kj} r_0(t)} \right) - \frac{\left(\beta_{kj}^2 - 1\right) v\mathrm{d}t}{2\sqrt{\beta_{kj}^2 r_0(t)^2 - (vt - r_0\cos\theta)^2}} + O(\mathrm{d}t^2) \overset{(a)}{=} \\
&\pi - \arccos\left(\frac{vt - r_0\cos\theta}{\beta_{kj} r_0(t)} \right) - \frac{\left(\beta_{kj}^2 - 1\right) v\mathrm{d}t}{2\sqrt{\beta_{kj}^2 r_0(t)^2 - (vt - r_0\cos\theta)^2}} + O(\mathrm{d}t^2),
\end{aligned}
\tag{C.7}
$$

where (a) follows from $\arccos(-x) = \pi - \arccos(x)$.

Using $r_0(t + \mathrm{d}t) = r_0(t)\left(1 + \frac{v\mathrm{d}t}{r_0(t)^2}(vt - r_0\cos\theta) + O(\mathrm{d}t^2)\right)$, when $\mathrm{d}t \to 0$, (C.6), and Taylor expansion of $(1+x)^{-1}$ also yields

$$
\begin{aligned}
&\arccos\left(\frac{vt - r_0\cos\theta}{\beta_{kj} r_0(t+\mathrm{d}t)} + \frac{\beta_{kj}^2 + 1}{2\beta_{kj}} \frac{v\mathrm{d}t}{r_0(t+\mathrm{d}t)} \right) = \\
&\arccos\left(\frac{vt - r_0\cos\theta}{\beta_{kj} r_0(t)} - \frac{(vt - r_0\cos\theta)^2}{r_0(t)^2} \frac{v\mathrm{d}t}{\beta_{kj} r_0(t)} + \frac{\beta_{kj}^2 + 1}{2\beta_{kj}} \frac{v\mathrm{d}t}{r_0(t)} + O(\mathrm{d}t^2) \right) = \\
&\arccos\left(\frac{vt - r_0\cos\theta}{\beta_{kj} r_0(t)} \right) - \frac{v\mathrm{d}t}{\sqrt{\beta_{kj}^2 r_0(t)^2 - (vt - r_0\cos\theta)^2}} \left(\frac{\beta_{kj}^2 - 1}{2} + 1 - \frac{(vt - r_0\cos\theta)^2}{r_0(t)^2} \right) + O(\mathrm{d}t^2).
\end{aligned}
$$

$$\tag{C.8}$$

From binomial series expansion, we also get

$$
\sqrt{\frac{2v\mathrm{d}t}{\beta_{kj}}\left(r_0(t)-\frac{vt-r_0\cos\theta}{\beta_{kj}}\right)+v^2\mathrm{d}t^2\left(1-\frac{1}{\beta_{kj}^2}\right)}\times
$$
$$
\sqrt{\frac{2v\mathrm{d}t}{\beta_{kj}}\left(r_0(t)+\frac{vt-r_0\cos\theta}{\beta_{kj}}\right)-v^2\mathrm{d}t^2\left(1-\frac{1}{\beta_{kj}^2}\right)}=\frac{2v\mathrm{d}t}{\beta_{kj}^2}\sqrt{\beta_{kj}^2 r_0(t)^2-(vt-r_0\cos\theta)^2}+O(\mathrm{d}t^2)
$$

$$(C.9)$$

Therefore, when $\mathrm{d}t\to 0$,

$$
\left|\mathcal{B}\left(\mathrm{x}(t),\frac{r_0(t)}{\beta_{kj}}\right)\setminus\mathcal{B}\left(\mathrm{x}(t+\mathrm{d}t),\frac{r_0(t+\mathrm{d}t)}{\beta_{kj}}\right)\right|=\pi\frac{r_0(t)^2}{\beta_{kj}^2}
$$
$$
-\left[\pi-\arccos\left(\frac{vt-r_0\cos\theta}{\beta_{kj}r_0(t)}\right)-\frac{\left(\beta_{kj}^2-1\right)v\mathrm{d}t}{2\sqrt{\beta_{kj}^2 r_0(t)^2-(vt-r_0\cos\theta)^2}}\right]\frac{r_0(t)^2}{\beta_{kj}^2}
$$
$$
-\left[\arccos\left(\frac{vt-r_0\cos\theta}{\beta_{kj}r_0(t)}\right)-\frac{v\mathrm{d}t}{\sqrt{\beta_{kj}^2 r_0(t)^2-(vt-r_0\cos\theta)^2}}\left(\frac{\beta_{kj}^2-1}{2}+1-\frac{(vt-r_0\cos\theta)^2}{r_0(t)^2}\right)\right]
$$
$$
\times\frac{r_0(t)^2+2v\mathrm{d}t(vt-r_0\cos\theta)}{\beta_{kj}^2}+\frac{v\mathrm{d}t}{\beta_{kj}^2}\sqrt{\beta_{kj}^2 r_0(t)^2-(vt-r_0\cos\theta)^2}+O(\mathrm{d}t^2),\qquad(C.10)
$$

where we have also used (5.17) for $\mathrm{d}t\to 0$. Finally, (5.19) can be obtained by simplifying (C.10).

C.3 Proof of Theorem 5.2

We can derive $\mathbb{E}[L\mid\text{tier}=k]$ using (5.12), i.e,

$$
\mathbb{E}[L\mid\text{tier}=k]=\lim_{z\to 0}\frac{z}{H_\ell(z\mid\text{tier}=k)}\overset{(a)}{=}\lim_{z\to 0}\frac{1}{\frac{\mathrm{d}}{\mathrm{d}z}H_\ell(z\mid\text{tier}=k)},\qquad(C.11)
$$

where (a) follows from L'Hospital's rule.

Since $|\mathcal{A}_{kj}(r_0,\theta,0,1,\beta_{kj})|=\pi\frac{r_0^2}{\beta_{kj}^2}$, we have

$$
\frac{\mathrm{d}}{\mathrm{d}z}H_\ell(z\mid\text{tier}=k)\Big|_{z=0}
$$
$$
=\frac{1}{\mathbb{P}(\text{tier}=k)}\int_0^\infty\int_0^\pi 2\lambda_k r_0\left(\sum_{j\in\mathcal{K}}\lambda_j\frac{\mathrm{d}}{\mathrm{d}z}|\mathcal{A}_{kj}(r_0,\theta,z,1,\beta_{kj})|\Big|_{z=0}\right)\exp\left\{-\sum_{j\in\mathcal{K}}\lambda_j\pi\beta_{jk}^2 r_0^2\right\}\mathrm{d}\theta\mathrm{d}r_0
$$
$$
=\frac{1}{\mathbb{P}(\text{tier}=k)}\int_0^\infty 2\lambda_k r_0\left(\sum_{j\in\mathcal{K}}\lambda_j\int_0^\pi\frac{\mathrm{d}}{\mathrm{d}z}|\mathcal{A}_{kj}(r_0,\theta,z,1,\beta_{kj})|\Big|_{z=0}\mathrm{d}\theta\right)\exp\left\{-\sum_{j\in\mathcal{K}}\lambda_j\pi\beta_{jk}^2 r_0^2\right\}\mathrm{d}r_0.
$$

(Note that in the above equations we have used $\beta_{jk} = \frac{1}{\beta_{kj}}$.)

By setting $z = 0$ in (5.28), we get

$$\frac{\mathrm{d}}{\mathrm{d}z} |\mathcal{A}_{kj}(r_0, \theta, z, 1, \beta_{kj})|\Big|_{z=0} =$$

$$\begin{cases} \frac{2r_0}{\beta_{kj}^2} \left[\sqrt{\beta_{kj}^2 - \cos^2 \theta} - \cos \theta \arccos \left(\frac{\cos \theta}{\beta_{kj}} \right) \right], \\ \qquad\qquad\qquad \text{if } (\beta_{kj} \geq 1) \text{ or } (\beta_{kj} < 1 \text{ and } \arccos(\beta_{kj}) < \theta < \pi - \arccos(\beta_{kj})) \\ 0, \qquad\qquad\qquad\qquad\qquad \text{if } (\beta_{kj} < 1 \text{ and } 0 \leq \theta \leq \arccos(\beta_{kj})) \\ -\frac{2\pi r_0 \cos \theta}{\beta_{kj}^2} \qquad\qquad\qquad \text{if } (\beta_{kj} < 1 \text{ and } \pi - \arccos(\beta_{kj}) \leq \theta \leq \pi) \end{cases}.$$

Therefore,

$$\int_0^\pi \frac{\mathrm{d}}{\mathrm{d}z} |\mathcal{A}_{kj}(r_0, \theta, z, 1, \beta_{kj})| \Big|_{z=0} \mathrm{d}\theta = 2r_0 \mathcal{I}(\beta_{kj}). \tag{C.13}$$

Substituting (C.13) in (C.12), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}z} H_\ell(z \mid \text{tier} = k)\Big|_{z=0} = \frac{\sum_{j \in \mathcal{K}} \lambda_j \mathcal{I}(\beta_{kj})}{\mathbb{P}(\text{tier} = k)} \int_0^\infty 4\lambda_k r_0^2 \exp \left\{ -\sum_{j \in \mathcal{K}} \lambda_j \pi \beta_{jk}^2 r_0^2 \right\} \mathrm{d}r_0$$

$$\overset{(a)}{=} \frac{\sum_{j \in \mathcal{K}} \lambda_j \mathcal{I}(\beta_{kj})}{\pi \left(\sum_{j \in \mathcal{K}} \lambda_j \beta_{jk}^2 \right)^{1/2}}, \tag{C.14}$$

where (a) follows from change of variable $\sum_{j \in \mathcal{K}} \lambda_j \pi \beta_{jk}^2 r_0^2 = t$. Finally, **Theorem 5.2** is derived by substituting (C.14) in (C.11).

Appendix D

Appendix to Chapter 6

D.1 Proof of Lemma 6.1

For device k at distance r_k from its associated BS, success probability is derived as follows

$$U_k = 1 - \mathbb{E}\left[\mathbf{1}\left(\max\left\{\text{SINR}_k(1), \text{SINR}_k(2), ..., \text{SINR}_k(\ell)\right\} < \theta\right) \mid k \in \mathcal{S}_t\right]$$

$$= 1 - \mathbb{E}_{\Phi_I}\left[\prod_{i=1}^{\ell} \mathbb{E}\left[\mathbf{1}\left(\text{SINR}_k(i) < \theta\right)\right] \mid k \in \mathcal{S}_t\right]$$

$$= 1 - \mathbb{E}_{\Phi_I}\left[\prod_{i=1}^{\ell} \mathbb{E}\left[\mathbf{1}\left(\frac{h_k(i)r_k^{-\alpha}}{I(i) + \sigma^2} < \theta\right)\right]\right]$$

$$\overset{(a)}{=} 1 - \mathbb{E}_{\Phi_I}\left[\left(1 - e^{-\theta\sigma^2 r_k^{\alpha}} \prod_{x \in \Phi_I} \frac{1}{1 + \theta r_k^{\alpha}\|x\|^{-\alpha}}\right)^{\ell}\right] \qquad (\text{D.1})$$

$$\overset{(b)}{=} \sum_{i=1}^{\ell} \binom{\ell}{i}(-1)^{i+1} e^{-i\theta\sigma^2 r_k^{\alpha}} \mathbb{E}_{\Phi_I}\left[\prod_{x \in \Phi_I} \frac{1}{(1 + \theta r_k^{\alpha}\|x\|^{-\alpha})^i}\right]$$

$$\overset{(c)}{\approx} \sum_{i=1}^{\ell} \binom{\ell}{i}(-1)^{i+1} \exp\left\{-i\theta\sigma^2 r_k^{\alpha} - \int_{\mathbb{R}^2}\left(1 - \frac{1}{(1 + \theta r_k^{\alpha}\|x\|^{-\alpha})^i}\right)\lambda(1 - e^{-12/5\lambda\pi\|x\|^2})\mathrm{d}x\right\}$$

$$\overset{(d)}{=} \sum_{i=1}^{\ell} \binom{\ell}{i}(-1)^{i+1} \exp\left\{-i\theta\sigma^2 r_k^{\alpha} - 2\pi\lambda\int_0^{\infty}\left(1 - \frac{1}{(1 + \theta r_k^{\alpha}x^{-\alpha})^i}\right)(1 - e^{-12/5\lambda\pi x^2})x\mathrm{d}x\right\},$$

where (a) is obtained by averaging with respect to the fading. (b) follows from using the binomial expansion. In reality, Φ_I is a subset of the set of the participating devices in the federated learning. However, since N is large in real world applications, selecting a device in each cell out of a massive number of devices for assigning a resource block at the beginning of each FL round is almost identical to uniformly selecting a point inside the cell. This model is referred to as *model of type I* in [22]. By studying the pair correlation

function (pcf) between a typical BS and interfering devices in this model, [22, 20] approximated Φ_I by a non-homogeneous PPP with intensity function $\lambda(x) = \lambda \left(1 - e^{-12/5\lambda\pi\|x\|^2} \right)$. In (c), we have used the probability generating functional (PGFL) for the point process Φ_I. Finally, (d) is obtained by using the polar domain representation.

Finally, note that averaging the success probability over r_k blurs the distinction between devices and removes the communication heterogeneity. In designing FL algorithms, however, we must incorporate the systems heterogeneity. Specifically, in FL algorithms, we must allocate higher weights to devices with lower success probabilities at the aggregation step which would not be possible if we averaged out r_k.

D.2 Proof of Theorem 6.1

We follow the same procedure as in [37]. Therefore, we mainly focus on the parts that are different.

When we choose the learning rate as $\eta_t = \frac{\beta}{\mu(\gamma+t)}$ with $\beta > 0$ and $\gamma = \max\left\{4\beta\frac{L}{\mu}, E\right\}$, it fulfills the following properties: i) η_t is decreasing with respect to t, ii) $\eta_t \leq \frac{1}{2L}$, $\forall t$, and iii) $\eta_t \leq 2\eta_{t+E}$, $\forall t$. These properties help us with the convergence analysis of the proposed federated learning algorithm.

In the following, we provide three lemmas which help us in proving the theorem.

Lemma D.1. *Averaging step in the proposed algorithm is unbiased, i.e.*

$$\mathbb{E}\left[\bar{w}_t\right] = \bar{v}_t, \qquad t \in \mathcal{I}_E,$$

where the expectation is over sampling and success event.

Proof. When $t \in \mathcal{I}_E$, $\bar{w}_t = w_t$; hence, from (6.8), we have

$$\mathbb{E}\left[\bar{w}_t\right] = \mathbb{E}\left[w_{t-E} + \sum_{k=1}^{N}\sum_{m=1}^{M}\frac{p_k}{q_k U_k}\mathbf{1}\left(k \in \mathcal{S}_{t-E}(m), \mathrm{SINR}_{k,m} > \theta\right)\left(v_t^k - w_{t-E}\right)\right]$$

$$\stackrel{(a)}{=} w_{t-E} + \sum_{k=1}^{N}\frac{p_k}{q_k U_k}\mathbb{E}\left[\sum_{m=1}^{M}\mathbf{1}\left(k \in \mathcal{S}_{t-E}(m)\right)\mathbb{E}\left[\mathbf{1}\left(\mathrm{SINR}_{k,m} > \theta\right) \mid k \in \mathcal{S}_{t-E}(m)\right]\right]\left(v_t^k - w_{t-E}\right)$$

$$= w_{t-E} + \sum_{k=1}^{N}\frac{p_k}{q_k U_k}\mathbb{E}\left[\sum_{m=1}^{M}\mathbf{1}\left(k \in \mathcal{S}_{t-E}(m)\right)U_k\right]\left(v_t^k - w_{t-E}\right)$$

$$= w_{t-E} + \sum_{k=1}^{N}\frac{p_k}{q_k U_k}q_k U_k(v_t^k - w_{t-E})$$

$$= \sum_{k=1}^{N}p_k v_t^k = \bar{v}_t,$$

where, in (a), the inner expectation is over the success event, and the outer expectation is with respect to the sampling. □

Lemma D.2. *When η_t satisfies the aforementioned properties,*

$$\mathbb{E}\left[\|\bar{w}_{t+1} - \bar{v}_{t+1}\|^2\right] \leq 4\eta_t^2 E^2 G^2 B, \qquad t+1 \in \mathcal{I}_E,$$

where $B = \sum_{k=1}^{N} p_k \left(\frac{1}{q_k U_k} - 1\right)$ for sampling Scheme I, and $B = \sum_{k=1}^{N} p_k \left(\frac{1}{q_k U_k} - \frac{1}{M}\right)$ for sampling Scheme II.

Proof. For brevity, in this proof, we denote the set of scheduled devices at time $t+1-E$ by $\mathcal{S}$ rather than $\mathcal{S}_{t+1-E}$.

$$\mathbb{E}\left[\|\bar{w}_{t+1} - \bar{v}_{t+1}\|^2\right]$$

$$= \mathbb{E}\left[\left\|w_{t+1-E} + \sum_{k=1}^{N}\sum_{m=1}^{M}\frac{p_k}{q_k U_k}\mathbf{1}\left(k \in \mathcal{S}(m), \text{SINR}_{k,m} > \theta\right)(v_{t+1}^k - w_{t+1-E}) - \sum_{k=1}^{N}p_k v_{t+1}^k\right\|^2\right]$$

$$= \mathbb{E}\left[\left\|\sum_{k=1}^{N}\sum_{m=1}^{M}\frac{p_k}{q_k U_k}\mathbf{1}\left(k \in \mathcal{S}(m), \text{SINR}_{k,m} > \theta\right)(v_{t+1}^k - w_{t+1-E}) - \sum_{k=1}^{N}p_k(v_{t+1}^k - w_{t+1-E})\right\|^2\right]$$

$$= \mathbb{E}\left[\left\|\sum_{k=1}^{N}p_k\frac{\sum_{m=1}^{M}\mathbf{1}\left(k \in \mathcal{S}(m), \text{SINR}_{k,m} > \theta\right) - q_k U_k}{q_k U_k}(v_{t+1}^k - w_{t+1-E})\right\|^2\right]$$

$$\stackrel{(a)}{\leq} \mathbb{E}\left[\sum_{k=1}^{N}p_k\left\|\frac{\sum_{m=1}^{M}\mathbf{1}\left(k \in \mathcal{S}(m), \text{SINR}_{k,m} > \theta\right) - q_k U_k}{q_k U_k}(v_{t+1}^k - w_{t+1-E})\right\|^2\right]$$

$$\stackrel{(b)}{=} \sum_{k=1}^{N}p_k\frac{\mathbb{E}\left[\left(\sum_{m=1}^{M}\mathbf{1}\left(k \in \mathcal{S}(m), \text{SINR}_{k,m} > \theta\right)\right)^2\right] - q_k^2 U_k^2}{q_k^2 U_k^2}\mathbb{E}\left[\|v_{t+1}^k - w_{t+1-E}\|^2\right], \qquad \text{(D.2)}$$

where (a) is obtained from convexity of $\|.\|^2$, and (b) follows from

$$\mathbb{E}\left[\sum_{m=1}^{M}\mathbf{1}\left(k \in \mathcal{S}(m), \text{SINR}_{k,m} > \theta\right)\right] = q_k U_k,$$

where the expectation is with respect to the sampling and success event. In the following, we first calculate $\mathbb{E}\left[\|v_{t+1}^k - w_{t+1-E}\|^2\right]$, and then calculate $\mathbb{E}\left[\left(\sum_{m=1}^{M}\mathbf{1}\left(k \in \mathcal{S}(m), \text{SINR}_{k,m} > \theta\right)\right)^2\right]$ for each sampling scheme separately.

$$\mathbb{E}\left[\|v_{t+1}^k - w_{t+1-E}\|^2\right] = \mathbb{E}\left[\left\|\sum_{i=t+1-E}^{t}\eta_i \nabla F_k(w_i^k; \xi_i^k)\right\|^2\right]$$

137

$$\overset{(a)}{\leq} \mathbb{E}\left[E \sum_{i=t+1-E}^{t} \left\|\eta_i \nabla F_k(w_i^k; \xi_i^k)\right\|^2\right]$$

$$\overset{(b)}{\leq} \mathbb{E}\left[\eta_{t+1-E}^2 E \sum_{i=t+1-E}^{t} \left\|\nabla F_k(w_i^k; \xi_i^k)\right\|^2\right]$$

$$\overset{(c)}{\leq} \eta_{t+1-E}^2 E^2 G^2 \overset{(d)}{\leq} 4\eta_{t+1}^2 E^2 G^2 \leq 4\eta_t^2 E^2 G^2, \tag{D.3}$$

where (a) is obtained by using Cauchy-Schwarz inequality. (b) follows from the property that η_t is decreasing with respect to t. In (c), we have used **Assumption 6.4**. (d) follows from $\eta_t \leq 2\eta_{t+E}$.

In *Scheme I*, at the aggregation step, each device can use at most one resource block. Thus,

$$\mathbb{E}\left[\left(\sum_{m=1}^{M} \mathbf{1}\left(k \in \mathcal{S}(m), \text{SINR}_{k,m} > \theta\right)\right)^2\right] = \mathbb{E}\left[\sum_{m=1}^{M} \mathbf{1}\left(k \in \mathcal{S}(m), \text{SINR}_{k,m} > \theta\right)\right] = q_k U_k. \tag{D.4}$$

However, in the second scheme, the BS may allocate more than one resource block to a device at a sampling time. $\mathbf{1}\left(k \in \mathcal{S}(m), \text{SINR}_{k,m} > \theta\right)$ is a Bernoulli random variable which takes value one with probability $\hat{q}_k U_k$. For *Scheme II*, sampling and success event over each resource block are i.i.d.; therefore, $\sum_{m=1}^{M} \mathbf{1}\left(k \in \mathcal{S}(m), \text{SINR}_{k,m} > \theta\right)$ is distributed according to a binomial distribution with parameters M and $\hat{q}_k U_k$, and we have

$$\mathbb{E}\left[\left(\sum_{m=1}^{M} \mathbf{1}\left(k \in \mathcal{S}(m), \text{SINR}_{k,m} > \theta\right)\right)^2\right] = q_k U_k \left(1 - \frac{1}{M} q_k U_k\right) + q_k^2 U_k^2$$

$$= q_k U_k + \left(1 - \frac{1}{M}\right) q_k^2 U_k^2. \tag{D.5}$$

$\square$

Lemma D.3. *When η_t satisfies the aforementioned properties, for any t, we have*

$$\mathbb{E}\left[\|\bar{v}_{t+1} - w^*\|^2\right] \leq (1 - \mu\eta_t)\mathbb{E}\left[\|\bar{w}_t - w^*\|^2\right] + \eta_t^2 \left(\sum_{k=1}^{N} p_k^2 \sigma_k^2 + 6L\Gamma + 8(E-1)^2 G^2\right).$$

Proof. See **Appendix A** in [37]. $\square$

Note that **Lemma D.1** and **Lemma D.2** only study the averaging steps ($t \in \mathcal{I}_E$), while **Lemma D.3** is a result of one step of SGD (for any t). **Lemma D.1** shows that each averaging step is unbiased, and **Lemma D.2** shows that variance of each averaging step is bounded. When, $t \notin \mathcal{I}_E$, $\bar{w}_t = \bar{v}_t$ according to (6.8).

Based on the above discussion, at any t, we have

$$
\begin{aligned}
\mathbb{E}\left[\|\bar{w}_{t+1} - w^*\|^2\right] &= \mathbb{E}\left[\|\bar{w}_{t+1} - \bar{v}_{t+1} + \bar{v}_{t+1} - w^*\|^2\right] \\
&\stackrel{(a)}{=} \mathbb{E}\left[\|\bar{w}_{t+1} - \bar{v}_{t+1}\|^2\right] + \mathbb{E}\left[\|\bar{v}_{t+1} - w^*\|^2\right] + 2\mathbb{E}\left[(\bar{w}_{t+1} - \bar{v}_{t+1})^T (\bar{v}_{t+1} - w^*)\right] \\
&\stackrel{(b)}{\leq} (1 - \mu\eta_t)\mathbb{E}\left[\|\bar{w}_t - w^*\|^2\right] + \eta_t^2\left(\sum_{k=1}^N p_k^2\sigma_k^2 + 6L\Gamma + 8(E-1)^2G^2 + 4E^2G^2B\right),
\end{aligned}
\tag{D.6}
$$

where the last term in (a) is zero based on **Lemma D.1** and the fact that $\bar{w}_t = \bar{v}_t$ when $t \notin \mathcal{I}_E$. In (b), we have used **Lemma D.2** and **Lemma D.3**.

For brevity, we define $C = \sum_{k=1}^N p_k^2\sigma_k^2 + 6L\Gamma + 8(E-1)^2G^2 + 4E^2G^2B$ and $\Delta_t = \|\bar{w}_t - w^*\|^2$. In the following, we find v such that $\mathbb{E}\left[\Delta_t\right] \leq \frac{v}{\gamma+t}$ at any t after initializing with Δ_0. This is satisfied at time $t = 0$ when $v \geq \gamma\Delta_0$. Moreover, when $\beta > 1$ and $v \geq \frac{\beta^2 C}{\mu^2(\beta-1)}$, $\mathbb{E}\left[\Delta_{t+1}\right] \leq \frac{v}{\gamma+t+1}$ given $\mathbb{E}\left[\Delta_t\right] \leq \frac{v}{\gamma+t}$. The proof is as follows:

$$
\begin{aligned}
\mathbb{E}\left[\Delta_{t+1}\right] &\leq (1 - \mu\eta_t)\mathbb{E}\left[\Delta_t\right] + \eta_t^2 C \\
&\leq \left(1 - \frac{\beta}{\gamma+t}\right)\frac{v}{\gamma+t} + \frac{\beta^2 C}{\mu^2(\gamma+t)^2} \\
&= \frac{\gamma+t-1}{(\gamma+t)^2}v + \left[\frac{\beta^2 C}{\mu^2(\gamma+t)^2} - \frac{\beta-1}{(\gamma+t)^2}v\right] \\
&\stackrel{(a)}{\leq} \frac{\gamma+t-1}{(\gamma+t)^2}v \leq \frac{v}{\gamma+t+1},
\end{aligned}
$$

where (a) is obtained from $v \geq \frac{\beta^2 C}{\mu^2(\beta-1)}$. Thus, by induction $\mathbb{E}\left[\Delta_t\right] \leq \frac{v}{\gamma+t}$ at any t when $v = \max\left\{\frac{\beta^2 C}{\mu^2(\beta-1)}, \gamma\Delta_0\right\}$.

Therefore, when $\eta_t = \frac{2}{\mu(\gamma+t)}$[1] with $\gamma = \max\{8\frac{L}{\mu}, E\}$, we have

$$
\mathbb{E}\left[\|\bar{w}_t - w^*\|^2\right] \leq \frac{\max\left\{\frac{4}{\mu^2}\left(\sum_{k=1}^N p_k^2\sigma_k^2 + 6L\Gamma + 8(E-1)^2G^2 + 4E^2G^2B\right), \gamma\|w_0 - w^*\|^2\right\}}{\gamma+t},
\tag{D.7}
$$

where $B = \sum_{k=1}^N p_k\left(\frac{1}{q_kU_k} - 1\right)$ for sampling *Scheme I*, and $B = \sum_{k=1}^N p_k\left(\frac{1}{q_kU_k} - \frac{1}{M}\right)$ for sampling Scheme II.

After the averaging step at time T ($T \in \mathcal{I}_E$), from L-smoothness of the global objective function F, we have

$$
\mathbb{E}\left[F(w_T) - F^*\right] \leq \frac{L}{2}\mathbb{E}\left[\|w_T - w^*\|^2\right]
$$

---

[1]We set $\beta = 2$.

$$\overset{(a)}{\leq} \frac{L}{2(\gamma+T)} \max\left\{ \frac{4}{\mu^2} \left(\sum_{k=1}^{N} p_k^2 \sigma_k^2 + 6L\Gamma + 8(E-1)^2 G^2 + 4E^2 G^2 B \right), \gamma \|w_0 - w^*\|^2 \right\}$$

$$\leq \frac{L/\mu}{\gamma+T} \left[\frac{2}{\mu} \left(\sum_{k=1}^{N} p_k^2 \sigma_k^2 + 6L\Gamma + 8(E-1)^2 G^2 + 4E^2 G^2 B \right) + \frac{\mu\gamma}{2} \|w_0 - w^*\|^2 \right],$$

where (a) is obtained from $w_T = \bar{w}_T$ and (D.7).

D.3   Proof of Section 6.4.3

With slight abuse of notation, we define $\hat{F}(w) = \sum_{k=1}^{N} \frac{p_k U_k}{\sum_{k'=1}^{N} p_{k'} U_{k'}} F_k(w)$, and we prove that (6.18) solves

$$\min_{w} \quad \hat{F}(w) = \sum_{k=1}^{N} \frac{p_k U_k}{\sum_{k'=1}^{N} p_{k'} U_{k'}} F_k(w), \tag{D.8}$$

which has the same solution as (6.19). Let us denote the solution to (D.8) by $\hat{w}^*$, and define $\alpha_k = p_k U_k$, $\alpha = \sum_{k=1}^{N} \alpha_k$, and $\alpha_k' = \frac{\alpha_k}{\alpha}$; therefore, $\sum_{k=1}^{N} \alpha_k' = 1$. For brevity, we also define $H_{k,m} = \mathbf{1}\left(k \in \mathcal{S}(m), \mathrm{SINR}_{k,m} > \theta\right)$, where we have ignored the time index of $\mathcal{S}$ since it is i.i.d. over different sampling steps.

Since we are using a different averaging approach, we must check **Lemma D.1** and **Lemma D.2**. However, **Lemma D.3** is not affected by the averaging steps; thus, it still holds (after replacing p_k with α_k'). We also need to change p_k to α_k' in definitions of $\bar{v}_t$, $\bar{w}_t$, g_t, and $\bar{g}_t$. It is worth reminding that, for sampling, we use *Scheme II* with $\hat{q}_k = p_k$.

Moreover, according to (6.18), when there is no successful transmission, the global model parameters at the BS do not change, which only affects the convergence rate (not the converging point). Therefore, to prove the convergence of (6.18) to $\hat{w}^*$, we assume $\sum_{k'=1}^{N} \sum_{m'=1}^{M} H_{k',m'} > 0$, i.e., at least one local update is available at the BS at each averaging step. Given $\sum_{k'=1}^{N} \sum_{m'=1}^{M} H_{k',m'} > 0$, we can write (6.18) as

$$w_t = \sum_{k=1}^{N} \sum_{m=1}^{M} \frac{H_{k,m}}{\sum_{k'=1}^{N} \sum_{m'=1}^{M} H_{k',m'}} v_t^k, \qquad t \in \mathcal{I}_E.$$

In the following, we provide a lemma that helps us derive **Lemma D.1** and **Lemma D.2** for the new averaging approach.

Lemma D.4. *For Scheme II sampling with $\{\hat{q}_k = p_k\}$, we have*

$$\mathbb{E}\left[\sum_{m=1}^{M} \frac{H_{k,m}}{\sum_{k'=1}^{N} \sum_{m'=1}^{M} H_{k',m'}} \mid \sum_{k'=1}^{N} \sum_{m'=1}^{M} H_{k',m'} > 0 \right] = \alpha_k'.$$

Proof.

$$\mathbb{E}\left[\sum_{m=1}^{M}\frac{H_{k,m}}{\sum_{k'=1}^{N}\sum_{m'=1}^{M}H_{k',m'}}\mid\sum_{k'=1}^{N}\sum_{m'=1}^{M}H_{k',m'}>0\right]$$

$$=\sum_{m=1}^{M}\mathbb{E}_{H_{k,m}}\left[\mathbb{E}\left[\frac{H_{k,m}}{\sum_{k'=1}^{N}\sum_{m'=1}^{M}H_{k',m'}}\mid\sum_{k'=1}^{N}\sum_{m'=1}^{M}H_{k',m'}>0,H_{k,m}\right]\right]$$

$$=\sum_{m=1}^{M}\mathbb{P}\left(H_{k,m}=1\mid\sum_{k'=1}^{N}\sum_{m'=1}^{M}H_{k',m'}>0\right)$$

$$\times\mathbb{E}\left[\frac{1}{1+\sum_{\substack{k'=1\\}}^{N}\sum_{\substack{m'=1,\\m'\neq m}}^{M}H_{k',m'}}\mid\sum_{k'=1}^{N}\sum_{m'=1}^{M}H_{k',m'}>0,H_{k,m}=1\right]$$

$$=\sum_{m=1}^{M}\sum_{i=0}^{M-1}\frac{1}{1+i}\mathbb{P}\left(\sum_{k'=1}^{N}\sum_{\substack{m'=1,\\m'\neq m}}^{M}H_{k',m'}=i,H_{k,m}=1\mid\sum_{k'=1}^{N}\sum_{m'=1}^{M}H_{k',m'}>0\right). \tag{D.9}$$

When $H_{k,m}=1$, we have $\sum_{k'=1}^{N}\sum_{m'=1}^{M}H_{k',m'}>0$. Moreover, resource allocation is i.i.d. over difference resource blocks, i.e. random variable $\sum_{k'=1}^{N}\sum_{\substack{m'=1,\\m'\neq m}}^{M}H_{k',m'}$ is independent of random variable $H_{k,m}$.

$$\mathbb{P}\left(\sum_{k'=1}^{N}\sum_{\substack{m'=1,\\m'\neq m}}^{M}H_{k',m'}=i,H_{k,m}=1\mid\sum_{k'=1}^{N}\sum_{m'=1}^{M}H_{k',m'}>0\right)$$

$$=\frac{\mathbb{P}\left(H_{k,m}=1\right)}{\mathbb{P}\left(\sum_{k'=1}^{N}\sum_{m'=1}^{M}H_{k',m'}>0\right)}\mathbb{P}\left(\sum_{k'=1}^{N}\sum_{\substack{m'=1,\\m'\neq m}}^{M}H_{k',m'}=i\right)$$

$$\stackrel{(a)}{=}\frac{\alpha_k}{1-(1-\alpha)^M}\binom{M-1}{i}\alpha^i(1-\alpha)^{M-1-i}, \tag{D.10}$$

where (a) is obtained using $\mathbb{P}\left(H_{k,m}=1\right)=\alpha_k$ and $\mathbb{P}\left(\sum_{k=1}^{N}H_{k,m}=1\right)=\alpha$. Finally, **Lemma D.4** is obtained by substituting (D.10) in (D.9). $\qquad\square$

When $t\in\mathcal{I}_E$, we have $\bar{w}_t=w_t$; thus,

$$\mathbb{E}\left[\bar{w}_t\mid\sum_{k'=1}^{N}\sum_{m'=1}^{M}H_{k',m'}>0\right]=\mathbb{E}\left[\sum_{k=1}^{N}\sum_{m=1}^{M}\frac{H_{k,m}}{\sum_{k'=1}^{N}\sum_{m'=1}^{M}H_{k',m'}}v_t^k\mid\sum_{k'=1}^{N}\sum_{m'=1}^{M}H_{k',m'}>0\right]$$

$$=\sum_{k=1}^{N}\mathbb{E}\left[\sum_{m=1}^{M}\frac{H_{k,m}}{\sum_{k'=1}^{N}\sum_{m'=1}^{M}H_{k',m'}}\mid\sum_{k'=1}^{N}\sum_{m'=1}^{M}H_{k',m'}>0\right]v_t^k$$

$$\stackrel{(a)}{=}\sum_{k=1}^{N}\alpha_k'v_t^k=\bar{v}_t, \tag{D.11}$$

141

where the expectation is with respect to sampling and success event and (a) follows from **Lemma D.4**.

At $t + 1 \in \mathcal{I}_E$, we also have

$$\mathbb{E}\left[\left\|\bar{w}_{t+1} - \bar{v}_{t+1}\right\|^2 \mid \sum_{k'=1}^N \sum_{m'=1}^M H_{k',m'} > 0\right]$$

$$= \mathbb{E}\left[\left\|\sum_{k=1}^N \frac{\sum_{m=1}^M H_{k,m}}{\sum_{k'=1}^N \sum_{m'=1}^M H_{k',m'}} \left(v_{t+1}^k - \bar{v}_{t+1}\right)\right\|^2 \mid \sum_{k'=1}^N \sum_{m'=1}^M H_{k',m'} > 0\right]$$

$$\leq \mathbb{E}\left[\sum_{k=1}^N \frac{\sum_{m=1}^M H_{k,m}}{\sum_{k'=1}^N \sum_{m'=1}^M H_{k',m'}} \left\|v_{t+1}^k - \bar{v}_{t+1}\right\|^2 \mid \sum_{k'=1}^N \sum_{m'=1}^M H_{k',m'} > 0\right]$$

$$\overset{(a)}{=} \mathbb{E}\left[\sum_{k=1}^N \alpha_k' \left\|v_{t+1}^k - \bar{v}_{t+1}\right\|^2\right] = \mathbb{E}\left[\sum_{k=1}^N \alpha_k' \left\|\left(v_{t+1}^k - w_{t+1-E}\right) - \left(\bar{v}_{t+1} - w_{t+1-E}\right)\right\|^2\right]$$

$$= \mathbb{E}\left[\sum_{k=1}^N \alpha_k' \left(\left\|v_{t+1}^k - w_{t+1-E}\right\|^2 + \left\|\bar{v}_{t+1} - w_{t+1-E}\right\|^2 - 2\left(v_{t+1}^k - w_{t+1-E}\right)^T \left(\bar{v}_{t+1} - w_{t+1-E}\right)\right)\right]$$

$$= \mathbb{E}\left[\sum_{k=1}^N \alpha_k' \left\|v_{t+1}^k - w_{t+1-E}\right\|^2\right] - \mathbb{E}\left[\left\|\bar{v}_{t+1} - w_{t+1-E}\right\|^2\right]$$

$$\leq \mathbb{E}\left[\sum_{k=1}^N \alpha_k' \left\|v_{t+1}^k - w_{t+1-E}\right\|^2\right] = \mathbb{E}\left[\sum_{k=1}^N \alpha_k' \left\|\sum_{i=t+1-E}^t \eta_i \nabla F_k(w_i^k, \xi_i^k)\right\|^2\right]$$

$$\leq \mathbb{E}\left[\sum_{k=1}^N \alpha_k' E \sum_{i=t+1-E}^t \left\|\eta_i \nabla F_k(w_i^k, \xi_i^k)\right\|^2\right] \leq \sum_{k=1}^N \alpha_k' E^2 \eta_{t+1-E}^2 G^2 \leq 4\eta_t^2 E^2 G^2, \tag{D.12}$$

where, to drive (a), we use **Lemma D.4**. The last line is also obtained similar to (D.3).

From (D.11), we understand that **Lemma D.1** holds (for problem (D.8)). Also, from (D.12), we understand that **Lemma D.2** holds with $B = 1$. As we discussed earlier, **Lemma D.3** is also valid (after replacing p_k with α_k'). Thus, following the same procedure as in **Appendix D.2**, we can prove that with using (6.18) at averaging steps, at time $T \in \mathcal{I}_E$, we have

$$\mathbb{E}\left[\left\|w_T - \hat{w}^*\right\|^2\right] \leq \frac{\max\left\{\frac{4}{\mu^2}\left(\sum_{k=1}^N \alpha_k'^2 \sigma_k^2 + 6L\Gamma + 8(E-1)^2 G^2 + 4E^2 G^2\right), \gamma \left\|w_0 - \hat{w}^*\right\|^2\right\}}{\gamma + T}.$$

Hence, the algorithm converges to the solution to (D.8).