The Dissertation Committee for Mandar Narsinh Kulkarni
certifies that this is the approved version of the following dissertation:

# System Design Issues in Dense Urban Millimeter Wave Cellular Networks

Committee:

Jeffrey G. Andrews, Supervisor

François Baccelli

Gustavo de Veciana

Robert W. Heath, Jr.

Sayandev Mukherjee

# System Design Issues in Dense Urban Millimeter Wave Cellular Networks

**by**

**Mandar Narsinh Kulkarni**

**DISSERTATION**

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**DOCTOR OF PHILOSOPHY**

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2018

Dedicated to the starry sky above me and the moral law within me.

(Philosopher Immanuel Kant's original quote)

# Acknowledgments

I big thank you to my supervisor Prof. Jeffrey G. Andrews for his guidance during my PhD. It has been a truly memorable journey working with him and I will always be inspired by his passion for high quality research. I will never be able to thank him enough for the time he invested in helping me grow as a researcher, the several opportunities he provided me and for the several insightful discussions, technical and otherwise, throughout last 5 years. I would also like to thank him for always being patient with me, encouraging me and understanding me during the ups and downs. I would also like to thank him for his classes on digital and wireless communications, which have helped me set up a solid foundation for research in wireless communications.

I would like to thank Prof. Robert W. Heath, Jr. for being a great inspiration for me through his several papers on millimeter wave wireless communication and for his classes on MIMO and wireless communications lab, which have been very useful to me. I would like to thank Prof. Gustavo de Veciana for several insightful discussions during the two classes I took with him on probability and queueing theory. It would not be wrong if I say those were among the best classes I have ever taken during my entire education. I would also like to thank Prof. Gustavo de Veciana for sharing with me ideas about deriving max-min rates on linear networks, which eventually helped me with

Chapter 5 in my thesis. I would like to thank Prof. François Baccelli for the classes on stochastic geometry and random graphs that I took with him. I will always try to follow a couple of thoughts he shared with me during some very enlightening discussions. First is about the attitude to find beauty in research and the second is a quote from Albert Einstein which he told me – the goal for mathematical models is to capture the essence of real world phenomenon in a simple way but not simplistic way. I would also like to thank Dr. Sayandev Mukherjee to graciously agree to serve on my dissertation committee.

I would like to thank Prof. Sanjay Shakkottai, Prof. Sujay Sanghavi, Prof. Alex Dimakis, Prof. Mihai Sirbu, and Prof. John Hasenbein who taught me classes at UT which added to my amazing experience during the PhD. I would like to thank Dr. Amitava Ghosh for hosting me at Nokia Bell Labs, Arlington Heights, IL during summers of 2014, 2015, and 2017. I would also like to thank Dr. Amitava Ghosh for supporting my studies, providing research ideas and giving regular feedback on my work. I would especially like to thank Dr. Tim Thomas, Dr. Fred Vook, Dr. Eugene Visotsky and Mr. Mark Cudak for their valuable collaboration during my internships. I would also like to thank Dr. Doru Calin and Dr. Aliye Ozge Kaya for hosting me at Nokia Bell Labs, Murray Hill, NJ in summer 2016. I would also like to thank Dr. Matthew Andrews from Nokia Bell Labs, Murray Hill, NJ for discussions regarding cross layer optimization in multi-hop wireless networks.

I would like to thank all my collaborators for their support. I would like to thank Abhishek Gupta, Abishek Sankararaman, Ahmed Alkhateeb, Arjun

Anand, Derya Malak, Murat Kocaoglu, Pablo Caballero, Rajat Sen, Sarabjot Singh, Xingqin Lin and Yingzhe Li for being inspiring mentors/colleagues during my PhD. I would like to thank all my colleagues at UT for various technical and fun discussions. I would especially like to thank Abishek Sankararaman, Aditya Mathad, Derya Malak, Erik Lindgren, Hardik Jain, Murat Kocaoglu, Preeti Kumari, Rajat Sen, Sara Mourad, Shalmali Joshi, Shaunak Kar, Soumya Basu, and Vivek Subramanian for helping me out during some tough times in May 2017. I would like to thank the International Office staff at UT as well as ECE staff for their help with the paperwork. I would like to thank my parents, sister, brother-in-law and niece for their constant love and support. I would also like to thank my relatives – grandparents, cousins, uncles, and aunts – for making my holidays back in India during the last 5 years special.

Chapter 2, Chapter 3, Chapter 4, and Chapter 5 are completely based on [1], [2], [3], and [4], respectively. The author of this dissertation is the primary author in [1–4]. The author thanks Dr. E. Visotsky for generating Fig. 3.1. The author thanks Manan Gupta for generating Fig. 5.2. The author also thanks Abishek Sankararaman, Arjun Anand and Soumya Basu for their feedback on Chapter 5, which significantly improved the mathematical rigour.

# System Design Issues in Dense Urban Millimeter Wave Cellular Networks

Publication No. _____

Mandar Narsinh Kulkarni, Ph.D.
The University of Texas at Austin, 2018

Supervisor: Jeffrey G. Andrews

Upcoming deployments of cellular networks will see an increasing use of millimeter wave (mmWave) frequencies, roughly between 20-100 GHz. The goal of this dissertation is to investigate some key design issues in dense urban mmWave cellular networks by developing mathematical models that are representative of these networks.

In the first contribution, stochastic geometry (SG) is used to study the per user rate performance of multi-user MIMO (MU-MIMO) in downlink mmWave cellular network incorporating the impact of a spatially sparse blockage dependent multipath channel and hybrid precoding. Performance of MU-MIMO is then compared with single-user beamforming and spatial multiplexing in different network scenarios considering coverage, rate and power consumption tradeoffs to suggest when to use which MIMO scheme.

The second contribution reconsiders a popular received signal power model used in system capacity analysis of MIMO wireless networks employ-

ing single user beamforming. A modification is suggested to the model by introducing a correction factor. An approximate analysis is done to justify incorporating such a factor and simulations are performed to validate it's importance. Although this contribution does not study a new system design issue for mmWave cellular, it highlights a shortcoming with using the popular received signal power model to study design issues in mmWave cellular networks.

The third and fourth contributions investigate resource allocation in self-backhauled mmWave cellular networks. In order to enable affordable initial deployments of mmWave cellular, self-backhauling is envisioned as a cost-saving solution. The third contribution investigates how to divide resources between uplink and downlink for access and backhaul in self-backhauled networks with single hop wireless backhauling. The performance of dynamic time division duplexing (TDD) and integrated access-backhaul (IAB) is compared with static TDD and orthogonal access backhaul (OAB) strategies using a SG based model. The last contribution of this dissertation addresses the following key question for self-backhauled networks. What is the maximum extended coverage area that a single fiber site can support using multi-hop relaying, while still achieving a minimum target per user data rate? The problem of maximizing minimum per user rates is studied considering a series of deployments with a single fiber site and varying number of relays. Several design guidelines for multi-hop mmWave cellular networks are provided based on the analytical and empirical results.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

"The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it." – Mark Weiser [6].

It is very hard to imagine a life without being able to use Google Maps while driving or to chat with someone at will using smart phones. Wireless networks have indeed become an integral part of our lives. Why did wireless networks become so successful? The answer lies in the ability of these networks to save time in our everyday life. These networks bring flexibility in getting a variety of things done without being constrained to be at a specific location to do so. Until the development of fourth generation of cellular networks, the primary focus was in enabling voice and primitive data communication amongst people. In the last decade the paradigm of Internet of Things [7], which is about establishing communication between several devices or belongings we have in a meaningful way, has picked up attention as well. Next generation wireless networks will also be about communication between people and devices, but these will differ from existing networks owing to the massive scale and diverse applications that will be supported.

Fifth generation (5G) of cellular networks will aim to support three major paradigms: (a) enhanced mobile broadband applications, (b) mission critical services and (c) massive internet of things [8, 9]. Enhanced mobile broadband applications, as the name suggests, will be about enabling applications that require very high data rates. For example, augmented or virtual reality applications. An example use-case for virtual/augmented reality will be to build toolsets for the purpose of designing and debugging new engineering systems [10–12]. Mission critical services will aim to support ultra reliable low latency applications [13,14]. For example, remote surgery or co-ordination amongst self-driving cars. Massive internet of things will aim to enable smart cities [15]. For example, massive deployment of smart meters or municipal garbage bins/street lights that automatically convey the need for attention to respective government authorities.

To enable enhanced mobile broadband applications, it is envisioned that the networks should support at least 100 Mbps per user data rates in 5G cellular networks, which is roughly 10-100 times more than current long term evolution (LTE) data rates [8,9,16]. Similarly, looking into the future if massive internet of things and mission critical services are also enabled supporting billions of devices, the current deployments of cellular networks are bound to be insufficient to support the multitude of new services, some requiring very high data rates. Following three ways or their combination are the primary means to meet this increasing demand in capacity – larger spectrum or extreme densification or higher spectral efficiency. The several Gigahertz (GHz) of

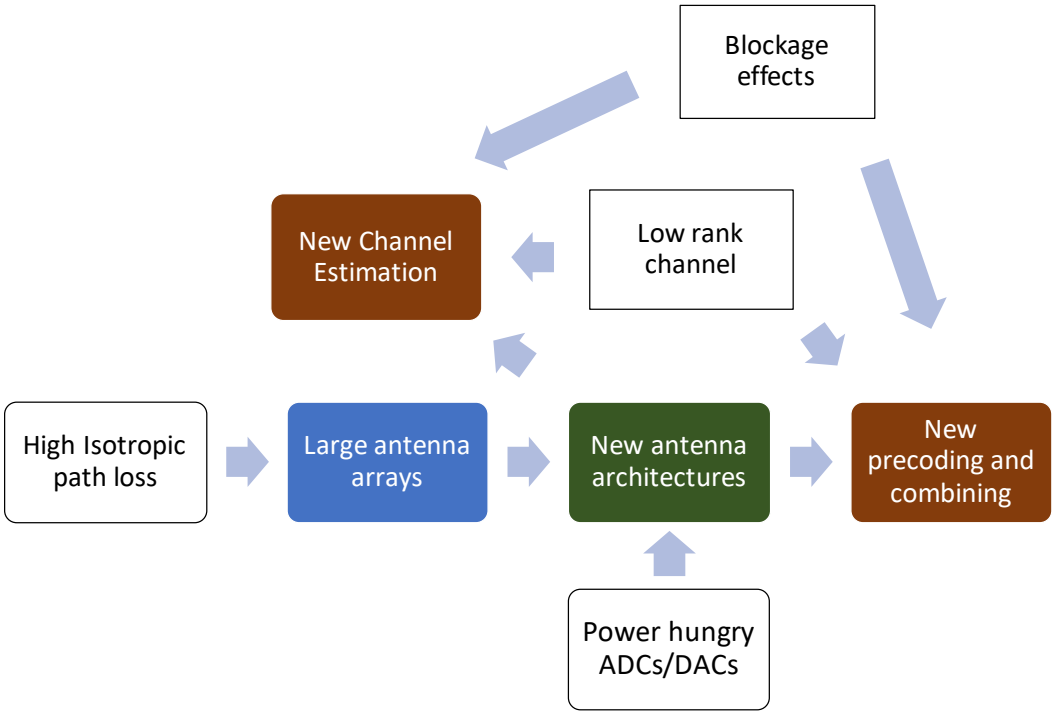



Figure 1.1: Physical layer disruptions in mmWave cellular networks.

underutilized spectrum at millimeter wave (mmWave) frequencies, between 30-100 GHz, makes it an attractive contender to design next generation cellular networks.

The objective of this thesis is to analytically study important system design issues in enabling mmWave cellular systems. The goal in all the contributions is to characterize achievable data rates as a function of different design choices keeping in mind mmWave specific propagation features and resulting impact on the hardware constraints for manufacturing the base stations or user equipments.

This introductory chapter has two sections. The first section will describe why cellular system design is challenging at mmWave frequencies. The second section will highlight the contributions of this dissertation in the light of the challenges discussed in the previous section.

## 1.1 Cellular System Design is Challenging at Millimeter Wave

In this section, some key disruptions in mmWave cellular system design are discussed motivated from distinct mmWave propagation and hardware constraints. Fig. 1.1 and Fig. 1.2 summarize some of the major disruptions and their corresponding motivations. These are described one by one in this section.

Near field path loss increases with square of carrier frequency. In order to mitigate the increased path loss, large antenna arrays that offer highly directional beams are believed to be a key enabling technology for mmWave cellular networks [17]. Note that this was identified as early as 1950s [18, 19] but the use of mmWave for cellular networks did not catch traction at that time, probably because of the availability of sufficient spectrum at lower frequencies which have much better propagation characteristics in line of sight (LOS) and non-LOS (NLOS) conditions. The requirement for directionality makes it challenging to perform initial access [20, 21] and also makes it necessary to perform beam tracking to enable mobility of user equipments [22]. The requirement of large antenna arrays also makes it necessary to develop

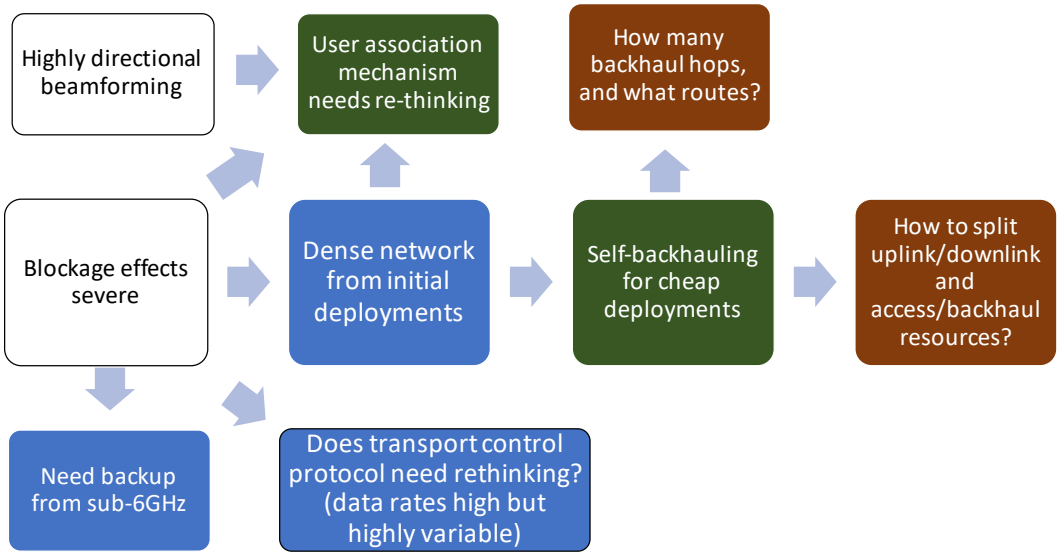



Figure 1.2: Network planning and MAC layer disruptions in mmWave cellular networks.

new signal processing techniques for channel estimation, precoding and combining to perform these operations with relatively low complexity [23]. Highly directional nature of mmWave transmissions lead to low rank nature of the mmWave channel [24, 25] and this is usually exploited to develop efficient signal processing algorithms [25–28]. Further, much larger available bandwidths – in the order of GHz contrasting against few tens of MHz for LTE – make fully digital precoding and combining impractical due to power hungry hardware components including analog to digital converters (ADCs), power amplifiers, etc. [23,29,30], and demands new antenna architectures like hybrid analog-digital beamformers or lens based analog front-ends or the use of few bit ADCs [23, 30, 31].

Blockage effects are notably enhanced at higher frequencies which causes the propagation loss to be highly sensitive to the blockage scenarios – presence of humans, trees, buildings in the surrounding cannot be neglected for performance evaluation [24,32–35]. This poses challenges in enabling seamless connectivity in environments wherein the blockages can be moving. Also, providing indoor coverage with outdoor deployment of mmWave cellular networks is a big challenge [33], although a recent ray tracing study has shown it may be feasible if indoor antennas are deployed close to windows [36]. Blockage effects makes it necessary to have a fall-back system with which the users can communicate if mmWave links fail due to blocking. It is envisioned that users will have dual connectivity with mmWave and sub-6GHz bands to avoid loss of important control signaling information due to mmWave link failures [21,36–38]. Another key implication of the significantly worse NLOS propagation as compared to LOS propagation is that mmWave networks need to have dense deployments right from initial phases to get sufficient coverage [21]. In order to enable low cost and flexible deployments, a promising strategy is to enable self-backhauling [39], meaning that a fraction of the base stations (BSs), called slave BSs (SBSs), wirelessly backhaul data to/from BSs with fiber backhaul, called master BSs (MBSs) sharing the time-frequency resources with the access links. This introduces new challenges at medium access control (MAC) layer as the SBSs may have to relay data to/from MBSs on multiple backhaul hops making it important to design efficient routing and scheduling algorithms in the context of mmWave cellular networks [**?**, 40–46].

It is clear from the above discussion that designing cellular networks on higher frequency bands is not a trivial extension of existing cellular networks, and needs significant rethinking of the physical and MAC layer issues. See [23, 47–49] for some detailed surveys on this topic. Recently, there has been an attempt to redesign transport control protocol (TCP) for mmWave networks as well [50]. The reason claimed is that very high data rates along with high variability can lead to high packet latencies with traditional TCP algorithms owing to the buffer-bloat phenomenon [51]. This dissertation investigates two design problems in dense urban mmWave cellular networks – (i) choice of MIMO techniques and (ii) finding theoretically optimal resource allocation schemes in multi-hop self-backhauled networks. The contributions of this thesis are described in detail in the next section.

## 1.2 Contributions and Organization

**Chapter 2. A Comparison of MIMO Techniques with Hybrid Beamforming in Downlink mmWave Cellular.** In this contribution, detailed in Chapter 2, a tractable model for rate in MU-MIMO mmWave cellular networks is proposed incorporating the two stage hybrid precoding algorithm in [52]. For the first time in stochastic geometry analysis of mmWave cellular networks, unlike previous works like [34, 46, 53], a spatially sparse blockage dependent channel model with rank greater than 1 is incorporated. Incorporating such a channel model is essential for fair comparison with single user spatial multiplexing. Virtual channel approximation [54] is used for quantifying the

zero forcing penalty in received signal power. The idea is to consider angles of arrival/departure to be taken from a quantized grid of size equal to number of antennas such that each possible array response vector is a column of a Discrete Fourier Transform (DFT) matrix. This gives an accurate handle on signal to noise ratio (SNR) distribution. For quantifying out of cell interference the virtual approximation, that leads to ON/OFF interference, is modified to include a notion of side-lobe gain and neglect the impact of zero forcing on interfering beams for tractability (this approximation is asymptotically tight with number of antennas). Upper and lower bounds to the Laplace transform of interference are derived and their accuracy is validated using Monte Carlo simulations.

Comparing the per user rates of MU-MIMO with single user beamforming (SU-BF) and spatial multiplexing (SM), it is found that MU-MIMO performs better in most practical operating scenarios assuming perfect channel state information is available. SM can outperform MU-MIMO in scenarios when we have sufficiently low user density coupled with reasonably large number of RF chains at user equipments (UEs)/BSs and multipath in the channel. However, due to the low rank nature of mmWave channels, this seems unlikely. Instead of fixing the density of BSs if power consumption per unit area is fixed, a denser SU-BF network outperforms MU-MIMO and SM in terms of per user cell edge rates. However, the sum rate with MU-MIMO is still usually better than SU-BF and SM. Incorporating the effect of possibly increased overheads with MU-MIMO/SM on achievable rates, the results in this chapter are re-

interpreted so as to quantify the minimum allowable efficiency for MU-MIMO to provide higher data rates than SM or SU-BF.

**Chapter 3. Correction Factor for Analysis of MIMO Wireless Networks With Highly Directional Beamforming.** In this chapter, a popular simplified received signal power model with single stream beamforming employed by a transmitter and a receiver is reconsidered in the regime when the beams have high gain and narrow beamwidth. *The correction factor* is defined as the ratio of the average actual received signal power divided by the average received signal power using the popular simplified model. This factor is analytically quantified for LOS and NLOS service and interfering links under some assumptions. The analysis along with simulations using a 3GPP compliant new radio (NR) channel model confirm the importance of incorporating the correction factor in coverage analysis of wireless networks that utilize the popular simplified received power model. Although this chapter does not study a new system design issue for mmWave cellular, it highlights a shortcoming of a popular modeling methodology for mmWave cellular and discusses implications on mmWave system design.

**Chapter 4. Resource Allocation in Self-backhauled Networks with a Single Backhaul Hop.** A fundamental problem for designing a self-backhauled network is to split the available time-frequency resources between uplink (UL) and downlink (DL) and for the access and backhaul links. In this contribution, detailed in Chapter 4, a random spatial model is developed for studying the coverage and mean rate performance of users in two hop

self-backhauled mmWave cellular networks as a function of different resource allocation schemes, with a focus on investigating the feasibility of dynamic time division duplexing (TDD) with unsynchronized access-backhaul (UAB) strategies. Definition of these schemes are given in Chapter 4. A two hop self-backhauled network implies that a user connects to a MBS or SBS. If it connects to a SBS, then the SBS directly communicates with an MBS without any SBS to SBS hops.

A time-slotted system is considered. In a typical access frame , all initial slots are prioritized for DL scheduling and later slots for UL scheduling. Such prioritization is shown to have inherent UL interference mitigation, which can be crucial for reducing outages with dynamic TDD. It is shown that low load and asymmetric UL-DL traffic are essential for gains with dynamic TDD. A switch between load aware static TDD and dynamic TDD would be desirable in high load interference-limited scenarios. Achievable mean rates with synchronized access-backhaul (SAB) and UAB are compared in self-backhauled mmWave cellular networks. It is found that there is no need for asymmetric traffic or low UE load for gains with UAB but we need sufficiently large number of SBS per MBS. The optimal number of slots to be exclusively allocated for access is observed to be non-increasing with UAB and dynamic TDD as compared to SAB and static TDD. With the model under consideration, self-backhauling is observed to be useful for enhancing coverage, but is not particularly useful to enhance mean rates if same antenna array is used by SBSs for both access and backhaul links. Employing higher spectral efficiency

backhaul links is important to reap benefits from dynamic TDD and UAB.

**Chapter 5. How Many Hops Can Self-backhauled Millimeter Wave Cellular Networks Support?** In the previous chapter, resource allocation in a self-backhauled network with single backhaul hop was studied. In this chapter, the following key question is addressed for designing financially viable mmWave cellular networks. What is the maximum extended coverage area that a single fiber site can support using multi-hop relaying, while still achieving a minimum target per user data rate? In order to answer this question, the maximum end to end data rate achieved by all users, called as max-min rates, is computed in a series of deployments with a single fiber site and different number of relays. A $k-$ring deployment model is proposed, wherein number of relays grows as $k^2$. Exact closed form expressions for max-min rates are derived considering integrated access backhaul (IAB) and orthogonal access-backhaul (OAB) schemes[1]. The analysis is also extended for full duplex relaying.

Several design guidelines for multi-hop mmWave cellular networks are given based on the analytical and empirical studies including choice of routing and scheduling strategy, maximum allowable self-interference in full duplex relays and impact of dual connectivity on system performance. For instance, it is shown that under realistic parameters for propagation losses and antenna gains, if BSs operating at 28 GHz with 800 MHz bandwidth are spaced at 100

[1]The terms IAB and UAB are used interchangeably in this dissertation. Similarly, SAB and OAB are used interchangeably.

meters with 2 worst case NLOS UEs each it is possible to support up to 6 rings of deployment serving 100 Mbps per UE, which covers an area of $1.2 \times 1.2$ km$^2$ if there is no upper limit on spectral efficiency that is converted to achievable data rates. For realistic values of upper limit on spectral efficiency, up to $k = 3$ can be supported for the parameters under consideration. Our empirical results indicate that mmWave self-backhauled networks in $k-$ring deployment can be noise-limited not only due to large bandwidth, narrow beamwidths and blockages but is also aided by the following reason. The proposed deployment has very few bottleneck links in several load scenarios considering reasonably large antenna gains (greater than 16 UE antennas and 64 BS antennas [16]) such that most NLOS UE access links are not bottlenecks. Thus, the optimal scheduler can meet the theoretically optimal max-min rates by just activating few links at a time, leading to noise-limited system performance. For routing, the optimality of nearest neighbour highway routing (defined in Section 5.4) is analytically proved in specific load scenarios, which is later observed to hold more generally through simulation studies.

# Chapter 2

# A Comparison of MIMO Techniques with Hybrid Beamforming[2]

A classical question in multi-antenna wireless communications has been to determine which MIMO technique performs better in different scenarios, for example based on the channel and interference characteristics. As mentioned in Section 1.1 at mmWave frequencies several important new factors must be considered, due to different hardware constraints on the precoders/combiners and a significantly different outdoor channel, which is both blockage-dependent and sparse (low rank) [24, 27, 55]. In order to compensate for the large near-field path loss, single user beamforming (SU-BF) has been the primary focus of several existing system capacity evaluations for mmWave cellular networks [17, 24, 34, 56]. However, recently there has been significant work on enabling multi-user MIMO (MU-MIMO) and single user spatial multiplexing (SM) under different antenna architectures that respect the necessary hard-

[2]This chapter reproduces the content of the following publication. M. N. Kulkarni, A. Ghosh and J. G. Andrews, "A comparison of MIMO techniques in downlink millimeter wave cellular networks", in *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1952-1967, May 2016. The research performed in this chapter including formulation of the problem, solving it and generating numerical results are primarily my contribution. My co-authors, A. Ghosh and J. G. Andrews, helped me in identifying the problem, giving regular feedback while I was working on the problem, and giving detailed feedback while I was writing the paper.

ware constraints at mmWave frequencies. Hybrid analog-digital precoders and combiners, receivers with low resolution analog to digital converters, and continuous aperture phase MIMO with lens-based beamformers (also called CAP-MIMO) are prominent antenna architectures being considered [30,57,58]. Most existing studies on mmWave MIMO, except for SU-BF, rely on single cell analysis for evaluating performance of the MIMO techniques and/or system level simulations for understanding the impact of base station (BS) deployment scenarios or blockages in the environment on the coverage and rate performance. Although analytical models for studying coverage and rate in SU-BF mmWave networks have been studied [34, 46], these cannot be directly used for studying other MIMO techniques like MU-MIMO and SM, as will be explained in Section 2.1.

The goals of this chapter are two-fold. First, a stochastic geometry-based model to study coverage and per user rate distribution in fully-connected hybrid beamforming-enabled MU-MIMO mmWave cellular networks is proposed. Second, this analytical model is used as a tool for comparing coverage, rate and power consumption for MU-MIMO, SM and SU-BF mmWave cellular networks.

## 2.1 Background and Related Work

Conventionally, BSs are equipped with fully-digital baseband processing. However, this approach requires a radio frequency (RF) chain per antenna which is impractical for mmWave BSs equipped with large antenna arrays.

Fully analog solutions, on the other hand, require only a single RF chain for the entire antenna array but have no capability of digital processing. Hybrid beamforming strikes a balance between these two solutions, wherein the number of RF chains can be designed to be between 1 (analog beamforming) and the number of antennas (digital beamforming). In a fully-connected architecture, each RF chain has phase shifters connected to all antennas in the array. On the other hand, in the array of sub-arrays architecture, the entire array is divided into sub-arrays and all antennas in a sub-array are connected via phase shifters to exactly one RF chain. The fully-connected architecture has higher beamforming gain than array of sub-arrays, for a fixed number of antennas. However, the power consumption and hardware complexity of precoder/combiner is lower in the latter. With low-complexity yet near optimal precoding/combining algorithms for MU-MIMO and SM being proposed with the fully-connected architecture [27, 52], this approach looks promising for practical implementation and is the focus of our discussion.

In [52], a joint baseband-RF precoder solution for MU-MIMO was proposed and proven to be asymptotically optimal as the number of antennas become large. Using this scheme, it was observed that MU-MIMO can offer higher sum rates than SU-BF. Another simulation-based work [59] highlighted that per user rates, including the cell edge rates, can be much higher with MU-MIMO with appropriate user pairing. It was observed that exploiting polarization diversity for two stream transmission to each user further enhances the gains in using MU-MIMO. This is one particular way in which SM gains

can be obtained in tandem with MU-MIMO. Another way to get SM gains would be to rely on the scatterers in the environment [27, 55]. The simulations in [27, 55] showed that SM and SU-BF could work in tandem to improve capacity. However, all these works implicitly neglected the aspect of power consumption at BSs and UEs when comparing the MIMO techniques. If we were to compare the coverage and rate performance of SU-BF and MU-MIMO or SM with fixed power consumption per unit area and fixed number of antennas per BS, we can deploy a much denser mmWave network with SU-BF than MU-MIMO or SM. This significantly affects the comparisons as will be shown in Section 2.7, since unlike in conventional cellular networks [60], densifying a mmWave network boosts the coverage and capacity notably [34, 56].

The above mentioned studies either rely on system-level simulations or on single cell analysis. There is no analytical model for MU-MIMO or SM mmWave networks that incorporates the impact of hybrid precoders and combiners and the channel sparsity. Analysis for MIMO cellular networks has conventionally been done by capturing the impact of linear precoding and combining into the distribution of an effective small scale fading random variable. In [61], it was shown that Gamma distribution can be used to model the small scale fading gain on serving and interfering links in MU-MIMO cellular networks employing ZF precoding. Most successive analytical studies on MU-MIMO cellular networks using stochastic geometry have relied on this result, for example [62, 63]. However, justifying this result assumes fully digital processing and full rank Rayleigh fading channel. At mmWave frequencies

the channel is expected to be sparse and blockage-dependent [24, 55, 64, 65]. Thus, the full rank assumption seems to be far from reality. A recent work [66] proposes an analytical model for SINR (signal to interference plus noise ratio) in MU-MIMO mmWave cellular networks but assumes fully digital processing. But, as described earlier, fully digital processing is also not realistic at mmWave. Analysis of multiuser mmWave cellular networks, thus, demands a new approach. Also, other existing analytical models for SU-BF enabled mmWave networks assume an equivalent SISO-like system with directional antenna gains by abstracting underlying signal level details [34, 46]. Further, the analysis in these papers is done for single path channels. An analytical framework that can be used as a tool for comparing with different MIMO techniques needs to incorporate multipath in the channel, which is a primary feature enabling SM. The key contributions in this work are as follows.

## 2.2 Contributions

### 2.2.1 Tractable Model for Coverage and Rate in MU-MIMO mmWave Cellular Networks

The analytical model captures the following mmWave-specific features: (i) precoding and combining with hybrid beamforming, and (ii) sparse blockage-depe ndent multipath channel model. For simplicity the channel model is assumed to be non-selective in both time and frequency to focus only on the spatial aspects. Using Monte-Carlo simulations, the model is shown to be reasonably accurate for a large number of antennas at the BSs and user equip-

ments (UEs) in noise-limited scenarios. In interference-limited scenarios, upper and lower bounds to the distribution of the proposed approximate SINR model are derived under some assumptions and validated with Monte-Carlo simulations. The fact that the proposed model incorporates different channel rank for line-of-sight (LOS) and non-LOS (NLOS) makes it possible to fairly compare analytical results with Monte-Carlo simulations for SM, which strongly depend on the rank of the channel. Numerical results reveal the following insights: (i) In interference-limited scenarios, SINR coverage has a non-monotonic trend with BS density. The optimum BS density for SINR coverage decreases with increasing degree of multiuser transmission. (ii) Although SINR coverage decreases with MU-MIMO, the median and peak per user rate increases due to increasing number of time slots available per user. However, the cell edge rates suffer with round robin scheduling, which motivates that the scheduler must explicitly safeguard the rates of edge users to use MU-MIMO.

### 2.2.2 Comparison of MIMO Techniques Considering Coverage, Rate and Power Consumption Tradeoffs

With perfect channel state information at the transmitter and neglecting channel acquisition and computational complexity overheads, MU-MIMO usually provides higher per user throughput compared to SM and SU-BF in mmWave networks for a fixed density of BSs and fixed number of antennas per BS/UE. Further note that enabling MU-MIMO requires only single RF chain at UEs, whereas enabling SM requires some baseband combining at UEs with multiple RF chains. A stochastic ordering argument is provided which

highlights that SNR coverage normalized by the antenna gains is better for MU-MIMO than SM asymptotically with the number of antennas at the BSs and users. SM can outperform MU-MIMO in scenarios when SM can support more streams than the number of users that can be served with MU-MIMO. This boils down to having sufficiently low user density coupled with sufficiently large number of RF chains at UEs/BSs and multipath in the channel. Instead of fixing the density of BSs if power consumption per unit area is fixed, a denser SU-BF network outperforms MU-MIMO and SM in terms of per user cell edge rates. However, the sum rate with MU-MIMO is still usually better than SU-BF and SM. The above results on sum or per user rates neglect the possibly increased overheads with MU-MIMO due to channel acquisition or computational complexity. Incorporating such factors, our results can be re-interpreted so as to quantify the minimum allowable efficiency for MU-MIMO to provide higher data rates than SM or SU-BF. The definition of minimum allowable efficiency is formally given in Section 2.5.2.

## 2.3 Organization and Notation

Section 2.4 sets up the system model. The analytical model for coverage and rate in MU-MIMO mmWave networks is developed in Section 2.5. Heuristic comparison of coverage and rate with SM is discussed in Section 2.6. Section 2.7 and 2.8 discusses the numerical results and conclusions.[3].

[3]Variables in italics are scalar random variables. Small and capital bold letters indicate vectors and matrices, respectively. An exception are random spatial locations in $\mathbb{R}^2$, which

## 2.4 System Model

Consider a downlink mmWave cellular network operating at carrier frequency $f_c$ and with bandwidth B. It is assumed that BSs and UEs are distributed in $\mathbb{R}^2$ as independent and homogeneous Poisson point processes (PPPs) $\Phi_{\text{BS}}$ and $\Phi_{\text{UE}}$, with intensities $\lambda_{\text{BS}}$ and $\lambda_{\text{UE}}$, respectively [60]. Each BS and user is assumed to employ a uniform linear array (ULA) of size $N_{\text{BS}}$ and $N_{\text{UE}}$, respectively. Full buffer traffic is assumed in this work.

### 2.4.1 Propagation Model

Path loss from BS at $x \in \Phi_{\text{BS}}$ to a user at $u \in \Phi_{\text{UE}}$ is given in dB by

$$L(x,u) = \beta + 10\alpha \log_{10}(||x-u||) + S_{x,u}, \tag{2.1}$$

where $\beta = 20\log_{10}\left(\frac{4\pi}{\lambda_c}\right)$ is the reference distance path loss at 1 meter, $\lambda_c$ is the wavelength in meters, $\alpha$ is the path loss exponent, $S_{x,u} \sim N(0,\xi^2)$ denotes Gaussian distribution with zero mean and standard deviation $\xi$. Note that $\alpha$ and $\xi$ are different for LOS and NLOS links. A subscript 'L' and 'N' to $\alpha$ and $\xi$ denote the respective parameters for LOS and NLOS links, respectively. A probabilistic blockage model proposed and validated in [46, 67] is used in this work. According to this model, the probability that a link of length $||x-u||$ is LOS is $p_{\text{LOS}}$ if $||x-u|| \leq$ D, for some value of D. All links longer than D are NLOS.

---

are italicized small letters $x, y, u, v$ or $w$. The complex conjugate transpose and pseudo inverse of $\mathbf{A}$ is $\mathbf{A}^*$ and $\mathbf{A}^\dagger$, respectively. Convergence in distribution is denoted by $\xrightarrow{d}$.

MmWave channels are expected to be sparse with very few angles of arrival (AOAs) and departure (AODs) capturing most of the energy [24, 55, 64, 65]. In this work, a narrowband geometric channel model [27, 52] is assumed, where the channel matrix between BS at $x$ and user $u$ is given by

$$\mathbf{H}_{x,u} = \sqrt{\frac{\mathrm{N_{BS}N_{UE}}}{L(x,u)\eta_{x,u}}} \sum_{i=1}^{\eta_{x,u}} \gamma_{i,x,u} \mathbf{a}_{\mathrm{UE}}(\phi_{i,x,u}) \mathbf{a}^*_{\mathrm{BS}}(\theta_{i,x,u}). \tag{2.2}$$

Here, $\eta_{x,u}$ is the number of paths between BS at $x$ and user at $u$, $\gamma_{i,x,u}$ is the small scale fading on $i^{\mathrm{th}}$ path (assumed to be complex normal with zero mean and unit variance for both LOS and NLOS to enhance analytical tractability), $\theta_{i,x,u}$ is the virtual AOD and $\phi_{i,x,u}$ is the virtual AOA for the $i^{\mathrm{th}}$ path. The number of paths $\eta_{x,u}$ equals $\eta_{\mathrm{L}}$ or $\eta_{\mathrm{N}}$ depending on whether the link is LOS or NLOS, respectively[4].

It is expected that $\eta_{\mathrm{N}} > \eta_{\mathrm{L}}$ [55, 64, 65]. The virtual AOA or AOD are related to the corresponding physical angles as $\theta = 2\pi \mathrm{d} \sin(\varphi)/\lambda_c$, where d is the inter-antenna spacing (chosen to be $\lambda_c/2$), $\varphi$ is the physical angle and $\theta$ is the virtual angle. The array response vectors for ULAs, $\mathbf{a}_{\mathrm{BS}}$ and $\mathbf{a}_{\mathrm{UE}}$, are of the form $\mathbf{a}(\theta) = [1\ e^{-j\theta} \dots e^{-j(\mathrm{N}-1)\theta}]^*/\sqrt{\mathrm{N}}$, where $\mathrm{N} \in \{\mathrm{N_{BS}}, \mathrm{N_{UE}}\}$. We assume that for every BS-UE link, scatterers in environment are uniformly distributed in $[0, 2\pi]$ and thus, the physical angles are also uniformly distributed in $[0, 2\pi]$.

[4] $\eta_{\mathrm{L}} > 1$ indicates more than 1 LOS *like* paths. In this work, there are either LOS or NLOS multipaths. A more general channel model would incorporate scenarios with 1 or more LOS like paths along with NLOS paths. However, an optimal power allocation would nearly allocate all power to LOS-like paths, thus, justifying our model. For simplicity, it is assumed that each scatterer gives rise to a single dominant path [52, 54, 68]. Extension to a clustered model [27, 54] is desirable in future.

This is called as the "physical model", which will be the basis of the simulation results in this chapter. However, for tractable analysis the virtual channel approximation [54] is leveraged in Section 2.5.2.

### 2.4.2 Fully Connected Hybrid Beamforming Architecture

A fully-connected two layer hybrid beamforming architecture with $\mathrm{N}_{\mathrm{RF}}^{\mathrm{BS}}$ and $\mathrm{N}_{\mathrm{RF}}^{\mathrm{UE}}$ RF chains at the BS and UE, respectively, is shown in Figure 2.1. A BS at $x$ sends a total of $\mathrm{N}_{\mathrm{s}}^{\mathrm{BS}}$ streams of data, which may include data sent to multiple users in the network. The transmit signals first go through a $\mathrm{N}_{\mathrm{RF}}^{\mathrm{BS}} \times \mathrm{N}_{\mathrm{s}}^{\mathrm{BS}}$ baseband precoder matrix $\mathbf{F}_x^{\mathrm{BB}} = [\mathbf{f}_{x,1}^{\mathrm{BB}} \dots \mathbf{f}_{x,\mathrm{N}_{\mathrm{s}}^{\mathrm{BS}}}^{\mathrm{BB}}]$ followed by a $\mathrm{N}_{\mathrm{BS}} \times \mathrm{N}_{\mathrm{RF}}^{\mathrm{BS}}$ RF precoder $\mathbf{F}_x^{\mathrm{RF}} = [\mathbf{f}_{x,1}^{\mathrm{RF}} \dots \mathbf{f}_{x,\mathrm{N}_{\mathrm{RF}}^{\mathrm{BS}}}^{\mathrm{RF}}]$. Note that the RF precoder is generally implemented using phase-shifters [27, 52], although there have been attempts trying to explore alternative methods [69]. Let us denote the RF combiner at user $u$ by $\mathbf{W}_u^{\mathrm{RF}}$ and the baseband combiner by $\mathbf{W}_u^{\mathrm{BB}} = [\mathbf{w}_{u,1}^{\mathrm{BB}}, \dots, \mathbf{w}_{u,\mathrm{N}_{\mathrm{s}}^{\mathrm{UE}}}^{\mathrm{BB}}]$. Note that SM, MU-MIMO and SU-BF can all be implemented with this generic architecture. The problem of jointly optimizing over $\mathbf{F}_x^{\mathrm{RF}}$, $\mathbf{F}_x^{\mathrm{BB}}$, $\mathbf{W}_u^{\mathrm{RF}}$ and $\mathbf{W}_u^{\mathrm{BB}}$ to maximize sum rate or per user rate for SM and MU-MIMO is still an open problem [27, 52]. In the following sections, the recently proposed near optimal algorithms for designing of precoders and combiners in [27] and [52] are assumed to employ SM and MU-MIMO, respectively as baseline for the simulations and analysis.

## 2.5 Multiuser MIMO in mmWave Cellular Networks

For MU-MIMO, it is assumed that each BS serves multiple users with a single stream per user. Thus, analog beamforming with a single RF chain suffices at each UE. Let $\mathcal{U}_x$ be the set of all users in $\Phi_{\mathrm{UE}}$ which are scheduled by the BS at $x$ in the same time slot, and the cardinality of $\mathcal{U}_x$ be $U_x$. We assume $U_x = \min(\mathrm{U_M}, N_x)$, where $N_x$ is the total number of users connected to the BS and $\mathrm{U_M}$ is the maximum number of users that can be scheduled in a time slot. A more sophisticated algorithm for deciding how many and which users to schedule in a resource block may be implemented as in [59, 70] but this aspect is neglected here for tractability. Furthermore, it is assumed that $\mathrm{U_M} = \mathrm{N_{RF}}$, and that unless the load on the BS is less than the number of RF chains, $\mathrm{U_M}$ users are served in a time slot. Also, when $\mathrm{U_M} > U_x$ only $U_x$ RF chains are used for processing, which means that $\mathbf{F}_x^{\mathrm{BB}} = [\mathbf{f}_{x,1}^{\mathrm{BB}} \dots \mathbf{f}_{x,U_x}^{\mathrm{BB}}]$ is of dimension $U_x \times U_x$ and $\mathbf{F}_x^{\mathrm{RF}} = [\mathbf{f}_{x,1}^{\mathrm{RF}} \dots \mathbf{f}_{x,U_x}^{\mathrm{RF}}]$ is of dimension $\mathrm{N_{BS}} \times U_x$.

Under the narrowband assumption, the received signal at user $u$ from BS at $x$ after passing through $\mathbf{w}_u$, the RF combiner at the user, is given by

$$y_u = \overline{\mathbf{h}}_{x,u}^{*} \mathbf{f}_{x,u}^{\mathrm{BB}} s_u + \sum_{v \in \mathcal{U}_x, v \neq u} \overline{\mathbf{h}}_{x,u}^{*} \mathbf{f}_{x,v}^{\mathrm{BB}} s_v + \mathrm{OCI} + \mathrm{noise},$$

where $\overline{\mathbf{h}}_{x,u}^{*} = \mathbf{w}_u^{*} \mathbf{H}_{x,u} \mathbf{F}_x^{\mathrm{RF}}$ and OCI is the out-of-cell interference. Here, $s_{(.)}$ are the transmit symbols with energy $\mathrm{P}/U_x$. Thus, the total transmit power of the BS is P. In this work, the precoding/combining algorithm in [52] is assumed considering an infinite resolution codebook at BSs and UEs for tractability. The first step is to design the RF precoders and combiners that maximize

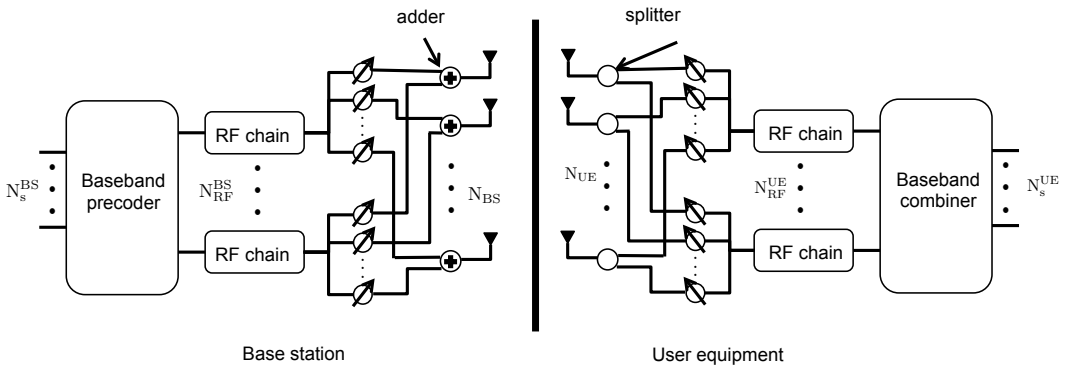



Figure 2.1: Fully-connected hybrid architecture at the BSs and UEs.

the received signal power on each of the BS-UE links. Thus, $\mathbf{w}_u$ and $\mathbf{f}_{x,u}^{\mathrm{RF}}$ are designed such that $(\mathbf{w}_u, \mathbf{f}_{x,u}^{\mathrm{RF}}) = \arg\max_{\mathbf{w},\,\mathbf{f}} |\mathbf{w}^* \mathbf{H}_{x,u} \mathbf{f}|$.

**Lemma 1** (from [68])**.** *The left and right singular vectors corresponding to non-zero singular values of* $\mathbf{H}_{x,u}$ *with* $\eta_{x,u} \ll \min(\mathrm{N_{BS}}, \mathrm{N_{UE}})$ *converge in chordal distance to* $\mathbf{a}_{\mathrm{UE}}(\phi_{i,x,u})$ *and* $\mathbf{a}_{\mathrm{BS}}(\theta_{i,x,u})$*, for* $1 \leq i \leq \eta_{x,u}$*. The corresponding singular values converge to* $\sqrt{\frac{\mathrm{N_{BS}N_{UE}}}{L(x,u)\eta_{x,u}}}|\gamma_{i,x,u}|$.

This lemma indicates that for large number of antennas $\mathbf{w}_u = \mathbf{a}_{\mathrm{UE}}(\phi_{i_m,x,u})$ and $\mathbf{f}_{x,u}^{\mathrm{RF}} = \mathbf{a}_{\mathrm{BS}}(\theta_{i_m,x,u})$, where $i_m = \arg\max_i |\gamma_{i,x,u}|$. This observation will be crucial in developing a tractable model for coverage and rate. Next, the baseband precoder is designed such that the multiuser interference is cancelled. Using a zero forcing (ZF) baseband precoder, $\mathbf{F}_x^{\mathrm{BB}} = \overline{\mathbf{H}}_x^{\dagger} \mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is a diagonal matrix whose entries are chosen such that $||\mathbf{F}_x^{\mathrm{RF}} \mathbf{f}_{x,u}^{\mathrm{BB}}|| = 1$. Here, $\overline{\mathbf{H}}_x = [\overline{\mathbf{h}}_{x,u_1} \dots \overline{\mathbf{h}}_{x,u_{U_x}}]^*$ with $\mathcal{U}_x = \{u_1, \dots, u_{U_x}\}$. Note that $\overline{\mathbf{H}}^{\dagger} = \overline{\mathbf{H}}^* \left(\overline{\mathbf{H}}\overline{\mathbf{H}}^*\right)^{-1}$, if $\overline{\mathbf{H}}$ is full rank.

### 2.5.1 SINR and Rate Model

The SINR of the user at $u \in \Phi_{\mathrm{UE}}$ served by a BS at $x \in \Phi_{\mathrm{BS}}$ connected to $U_x$ total users is given by

$$\mathsf{SINR}_{x,u} = \frac{\frac{||\overline{\mathbf{h}}_{x,u}^{*}\mathbf{f}_{x,u}^{\mathrm{BB}}||^2}{U_x}}{\frac{\sigma_n^2}{\mathrm{P}} + \sum\limits_{\substack{v\in\mathcal{U}_x \\ v\neq u}} \frac{||\overline{\mathbf{h}}_{x,u}^{*}\mathbf{f}_{x,v}^{\mathrm{BB}}||^2}{U_x} + \sum\limits_{\substack{y\in\Phi_{\mathrm{BS}} \\ y\neq x}} \sum\limits_{w\in\mathcal{U}_y} \frac{||\overline{\mathbf{h}}_{y,u}^{*}\mathbf{f}_{y,w}^{\mathrm{BB}}||^2}{U_y}}. \tag{2.3}$$

The second term in the denominator is zero, owing to the ZF precoder and the fact that $\overline{\mathbf{H}}_x$ is almost surely full rank for independently distributed channel gains from BS at $x$ to users in $\mathcal{U}_x$. The per user rate (in bits per second or bps) of user $u$ served by BS at $x$ is defined as

$$R_{x,u} = \omega_x \frac{\mathrm{B}\, U_x}{N_x} \log_2(1 + \mathsf{SINR}_{x,u}), \tag{2.4}$$

where $\omega_x < 1$ models the efficiency in implementing MU-MIMO in terms channel acquisition or computational complexity or cyclic prefix while implementing OFDM [66, 71]. The above model implies that each user gets $U_x/N_x$ fraction of resources, which can be achieved using round robin scheduling. The sum rate is defined as

$$R_x = \omega_x \mathrm{B} \sum_{u\in\mathcal{U}_x} \log_2(1 + \mathsf{SINR}_{x,u}), \tag{2.5}$$

which is basically the total number of bits per second (bps) transmitted by the BS, whereas the per user rate is the rate achieved by a typical user in a scheduling cycle.

In general, the efficiency factors vary for different BSs and are dependent on $\mathrm{U_M}, \mathrm{N_{BS}}, \mathrm{N_{UE}}$,$\eta_{\mathrm{N}}$, $\eta_{\mathrm{L}}$ and OFDM cyclic prefix penalty. For simplicity,

it is assumed that $\omega_x = \omega_{\mathrm{MU}}, \forall x \in \Phi_{\mathrm{BS}}$. One can interpret $\omega_{\mathrm{MU}} = \min_x \omega_x$ to get a lower bound on the rate. Since the overhead is expected to increase with $U_x$, $\omega_{\mathrm{MU}}$ corresponds to the efficiency of BSs serving $\mathrm{U_M}$ users. Note that $\omega_{\mathrm{MU}}$ for $\mathrm{U_M} = 1$ is the overhead for SU-BF.

### 2.5.2 Coverage and Rate Analysis

Consider a *typical* UE at origin, wherein the notion of typicality for stationary point process is defined through Palm probability [72], and it associates with the BS at $x$ offering minimum path loss $L(x, 0)$. We call this the *tagged* BS. We evaluate the $\mathsf{SINR}$ coverage defined as $\mathbb{P}\left(\mathsf{SINR}_{x,0} > \tau\right)$, which is the $\mathsf{SINR}$ distribution of the *typical* user at origin. Rate coverage is similarly defined. The $\mathsf{SINR}$ expression in (2.3), although exact, is not tractable in terms of finding its distribution. We, thus, provide an accurate yet tractable approximation that captures the dependency of the several parameters in the following analysis.

**Definition 1.** A random variable $\mathrm{Z}_1$ stochastically dominates another random variable $\mathrm{Z}_2$, if $\mathbb{P}(Z_1 > z) \geq \mathbb{P}(Z_2 > z)$ for all $z \in \mathbb{R}$. We denote this as $Z_1 \overset{\mathrm{st}}{\geq} Z_2$.

#### 2.5.2.1 Rate Distribution in a Noise-limited Network

We first focus on finding the rate distribution in a network with negligible interference effects. Throughout the discussion, the virtual angles of departure/arrival are quantized to take values in $\{\theta : \theta = -\pi + \frac{2\pi i}{\mathrm{N}_a}, 1 \leq i \leq \mathrm{N}_a\}$.

**Lemma 2.** *If antenna spacing is half wavelength and the physical AOAs/AODs*

*are uniformly distributed in 0 to $2\pi$, the distribution of the quantized virtual angles is given by*

$$q_{a,i} = \frac{\left(\sin^{-1}\left(-1+\frac{2i+1}{\mathrm{N}_a}\right) - \sin^{-1}\left(-1+\frac{2i-1}{\mathrm{N}_a}\right)\right)}{\pi},$$

*for $a \in \{\mathrm{UE}, \mathrm{BS}\}$ and $i \in \{1, \ldots, \mathrm{N}_a - 1\}$. Further, $q_{a,\mathrm{N}_a} = 1 - \sum_{j=1}^{\mathrm{N}_a - 1} q_{a,j}$.*

*Proof.* Note that $\theta = \pi \sin(\varphi)$ for half wavelength antenna spacing. Thus, the required probability can be computed by using that $\varphi$ is uniformly distributed in 0 to $2\pi$. □

**Proposition 1.** *Let $\mathcal{U}_x = \{u_1, \ldots, u_{\mathrm{U_x}}\}$ be the users served by the BS at $x$. Assuming $\eta_N \ll \min(\mathrm{N_{BS}}, \mathrm{N_{UE}})$, $\mathrm{U_M} \ll \min(\mathrm{N_{BS}}, \mathrm{N_{UE}})$ and a dense network deployment, $\mathsf{SNR}$ at user $u_1$ can be modelled as*

$$\mathsf{SNR}_{x,u_1} \approx \frac{\mathrm{G}}{\eta_{x,u_1} U_x \sigma_n^2} |\gamma_{i_m,x,u_1}|^2 L(x,u_1)^{-1} p_{\mathrm{ZF}}, \tag{2.6}$$

*where $\mathrm{G} = \mathrm{P}\mathrm{N_{BS}}\mathrm{N_{UE}}$, $i_m$ is the index corresponding to $\arg\max_i |\gamma_{i,x,u_1}|$, $p_{\mathrm{ZF}}$ is a random variable that captures reduction in signal power due to the ZF penalty and has distribution that stochastically dominates $p_{\mathrm{MU}}$, which is a Bernoulli random variable with success probability $\zeta(\eta_{x,u_1}, U_x)$, where*

$$\zeta(\eta_{x,u_1}, U_x) = \sum_{j=1}^{\mathrm{N_{BS}}} q_{\mathrm{BS},j} \mathrm{B}_j(\eta_{x,u_1}, U_x) \left(\mathrm{p_{LOS}} \mathrm{A}_j(\eta_{\mathrm{L}}) + (1 - \mathrm{p_{LOS}}) \mathrm{A}_j(\eta_{\mathrm{N}})\right)^{U_x - 1},$$

$\mathrm{A}_j(\eta) = \sum_{i=1}^{\mathrm{N_{UE}}} q_{\mathrm{UE},i} \left(1 - q_{\mathrm{UE},i} q_{\mathrm{BS},j}\right)^{\eta - 1}$, $\mathrm{B}_j(\eta, , U_x) = \mathrm{C}(\eta)(1 - q_{\mathrm{BS},j})^{U_x - 1} + \mathrm{D}_j(\eta, U_x) - \mathrm{D}_j(\eta, U_x)\mathrm{C}(\eta)$, $\mathrm{C}(\eta) = \sum_{i=1}^{\mathrm{N_{UE}}} q_{\mathrm{UE},i}(1 - q_{\mathrm{UE},i})^{\eta - 1}$, *and*

$$\mathrm{D}_j(\eta, U_x) = \sum_{i_1, \ldots, i_{\eta-1}=1}^{\mathrm{N_{BS}}-1} \prod_{n=1}^{\eta-1} l_{i_n,j} \left(1 - q_{\mathrm{BS},j} - \sum_{unique(i_{(.)})} l_{i_n,j}\right)^{U_x - 1},$$

*where* $l_{n,j} = q_{\mathrm{BS},n}$ *if* $n < j$ *and* $l_{n,j} = q_{\mathrm{BS},n+1}$ *if* $n \geq j$ *and unique*$(i_{(.)})$ *represents the unique values in the set* $\{i_1, \ldots, i_{\eta-1}\}$.

*Proof.* Without loss of generality, $i_m = 1$. From Lemma 1, $\mathbf{w}_u = \mathbf{a}_{\mathrm{UE}}(\phi_{1,x,u})$ and $\mathbf{f}_{x,u}^{\mathrm{RF}} = \mathbf{a}_{\mathrm{BS}}(\theta_{1,x,u})$. Using the orthogonality of the array response vectors with quantized virtual angles, it can be shown that $\overline{\mathbf{H}}_x$ takes the form

$$\overline{\mathbf{H}}_x = \begin{bmatrix} \sqrt{\frac{\mathrm{N_{BS}N_{UE}}}{L(x,u_1)\eta_{x,u_1}}}\gamma_{1,x,u_1} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{P}}_x \end{bmatrix}, \tag{2.7}$$

with probability at least $\zeta(\eta_{x,u_1}, U_x)$. See Appendix 2.9.1 for details. Note that here $\tilde{\mathbf{P}}_x$ is a submatrix of $\overline{\mathbf{H}}_x$ of dimension $U_x - 1 \times U_x - 1$. In this case,

$$\overline{\mathbf{H}}_x^{\dagger} = \begin{bmatrix} \sqrt{\frac{L(x,u_1)\eta_{x,u_1}}{\mathrm{N_{BS}N_{UE}}}}\gamma_{1,x,u_1}^{-1} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{P}}_x^{\dagger} \end{bmatrix}.$$

We know that $\mathbf{F}_x^{\mathrm{BB}} = \overline{\mathbf{H}}_x^{\dagger}\mathbf{\Lambda}$, for diagonal matrix $\mathbf{\Lambda}$ that helps satisfy the power constraints. Thus, the first column of the baseband precoder is of the form $\mathbf{f}_{x,u_1}^{\mathrm{BB}} = [c\ 0 \ldots 0]$, for some constant $c$ such that $||\mathbf{F}_x^{\mathrm{RF}}\mathbf{f}_{x,u_1}^{\mathrm{BB}}|| = 1$. Thus, $\mathbf{f}_{x,u_1}^{\mathrm{BB}} = [1\ 0 \ldots 0]$ since each term in $\mathbf{F}_x^{\mathrm{RF}}$ is unit norm. In this case, the received signal power of $u_1$ is equal to $\frac{G}{\eta_{x,0}U_x}|\gamma_{1,x,u_1}|^2 L(x,u_1)^{-1}$, which corresponds to the case when $p_{\mathrm{MU}} = 1$ in (2.6). Since the event that $\mathbf{P}_x$ is not of this form is of low probability and results in even more complex expressions, the signal power is lower bounded by 0 in this case. Under virtual channel approximation, (2.6) is a lower bound on $\mathsf{SNR}$. □

**Remark 1.** *If the quantized virtual angles are distributed uniformly in their range, instead of the distribution in Lemma 2,* $\mathrm{D}_j(\eta, U_x)$ *takes a much sim-*

*plified form given by* $\mathrm{D}(\eta, U_x) = \sum_{d=1}^{\eta-1} \binom{\mathrm{N_{BS}}-1}{d} \frac{(\mathrm{N_{BS}}-1-d)^{U_x-1}}{(\mathrm{N_{BS}}-1)^{U_x+\eta-2}} \sum_{i=0}^{d} (-1)^i (d-i)^{\eta-1} \binom{d}{i}$.

**Remark 2.** *To simplify evaluation of Proposition 1, the following can be used*

$$(1-q_{\mathrm{BS},j})^{U_x-1} \sum_{i=1}^{\mathrm{N_{UE}}} q_{\mathrm{UE},i}(1-q_{\mathrm{UE},i})^{\eta-1} \leq \mathrm{B}_j(\eta, U_x) \leq (1-q_{\mathrm{BS},j})^{U_x-1}.$$

**Remark 3.** *It can be shown that* $\sum_{i=1}^{\mathrm{N_{BS}}} q_{\mathrm{BS},i}(1-q_{\mathrm{BS},i})^r \to 1$ *as* $\mathrm{N_{BS}} \to \infty$ *for any* $r \geq 0$*, which is true since* $\max_i q_{\mathrm{BS},i} \to 0$ *as* $\mathrm{N_{BS}} \to \infty$*. Similar result holds for* $q_{\mathrm{UE},j}$ *with* $\mathrm{N_{UE}} \to \infty$*. All these imply that* $\zeta \to 1$ *with* $\mathrm{N_{BS}} \to \infty$ *and* $\mathrm{N_{UE}} \to \infty$*.*

To find the SNR coverage, the distribution of the path loss and number of multiuser streams in Proposition 1 needs to be found. First the focus will be on the finding the probability mass function (PMF) of the number of multiuser streams of BS at $y \in \Phi_{\mathrm{BS}}$ given by $U_y = \min\{\mathrm{U_M}, N_y\}$. An approximation proposed in [73] is used to model the distribution of $N_y$, which are actually correlated random variables for $y \in \Phi_{\mathrm{BS}}$ and particularly known to be intractable since finding the volume of Voronoi association cells is itself an unsolved problem [74]. With notably different propagation channels for LOS and NLOS links, the cell association regions in mmWave networks are not even Voronoi and more irregular [46]. The PMF of $N_y$ is denoted by $\kappa(n)$ is modelled as follows [46]. Let $\rho = \lambda_{\mathrm{UE}}/\lambda_{\mathrm{BS}}$, then if $y = x$, that is the BS is serving the typical user, $\kappa(n)$ is approximated by

$$\frac{3.5^{3.5}}{(n-1)!} \frac{\Gamma(n+3.5)}{\Gamma(3.5)} \rho^{n-1} (3.5+\rho)^{-n-3.5}, \tag{2.8}$$

for $n \geq 1$ and $\kappa(0) = 0$. For interfering BSs, $\kappa(n) =$

$$\frac{3.5^{3.5}}{n!}\frac{\Gamma(n+3.5)}{\Gamma(3.5)}\rho^n\left(3.5+\rho\right)^{-n-3.5}, \tag{2.9}$$

for $n \geq 0$.

Assuming $N_y$ to be i.i.d., the PMF of $U_y$ is modeled as

$$\mathbb{P}\left(U_y = n\right) = \mathbb{1}_{\{0\leq n\leq \mathrm{U_M}-1\}}\kappa(n) + \left(1 - \sum_{i=1}^{\mathrm{U_M}-1}\kappa(n)\right)\mathbb{1}_{\{n=\mathrm{U_M}\}}. \tag{2.10}$$

To find the path loss distribution, which is blockage dependent, the point process $\Phi_{\mathrm{BS}}$ is modeled to be superposition of the point processes $\Phi_{\mathrm{L}}$ and $\Phi_{\mathrm{N}}$ with intensities $\lambda_{\mathrm{BS}}p_{\mathrm{LOS}}\mathbb{1}_{\{||x||\leq \mathrm{D}\}}$ and $\lambda_{\mathrm{BS}}(1-p_{\mathrm{LOS}})\mathbb{1}_{\{||x||\leq \mathrm{D}\}}+\lambda_{\mathrm{BS}}\mathbb{1}_{\{||x||>\mathrm{D}\}}$, respectively. These two point processes correspond to LOS and NLOS BSs. The corresponding propagation processes [75] are given as $\mathcal{N}_{\mathrm{L}} = \{||y||^{\alpha_{\mathrm{L}}}/S_{y,\mathrm{L}} : y \in \Phi_{\mathrm{L}}\}$, and $\mathcal{N}_{\mathrm{N}} = \{||y||^{\alpha_{\mathrm{N}}}/S_{y,\mathrm{N}} : y \in \Phi_{\mathrm{N}}\}$.

**Lemma 3.** *$\mathcal{N}_{\mathrm{L}}$ is a non-homogeneous PPP with intensity $\Lambda_{\mathrm{L}}([0,t)) = \lambda_{\mathrm{BS}}\mathrm{M_L}(t)$, where $\mathrm{M_L}(t)$ is given as follows.*

$$\mathrm{M_L}(t) = \pi p_{\mathrm{LOS}}\left[\mathrm{D}^2\mathsf{Q}\left(\Upsilon_{\mathrm{L}}(t)\right) + t^{\frac{2}{\alpha_{\mathrm{L}}}}\exp\left(\frac{2\sigma_{\mathrm{L}}^2}{\alpha_{\mathrm{L}}^2}+\frac{2m}{\alpha_{\mathrm{L}}}\right)\mathsf{Q}\left(\frac{2\sigma_{\mathrm{L}}^2}{\alpha_{\mathrm{L}}\sigma_{\mathrm{L}}} - \Upsilon_{\mathrm{L}}(t)\right)\right].$$

*Here, $m = -0.1\beta\ln 10$, $\sigma_{\mathrm{L}} = 0.1\xi_{\mathrm{L}}\ln 10$, $\Upsilon_j(t) = \frac{\ln(\frac{\mathrm{D}^{\alpha_j}}{t})-m}{\sigma_j}$ for $j \in \{\mathrm{L},\mathrm{N}\}$ and $\mathsf{Q}(.)$ is the Q-function (Standard Gaussian CCDF).*

*Proof.* Special case of Appendix A of [46] and is therefore skipped for brevity. □

**Lemma 4.** *$\mathcal{N}_{\mathrm{N}}$ is a non-homogeneous PPP with intensity $\Lambda_{\mathrm{N}}([0,t)) = \lambda_{\mathrm{BS}}\mathrm{M}_{\mathrm{N}}(t)$, where $\mathrm{M}_{\mathrm{N}}(t)$ is given as follows.*

$$\mathrm{M}_{\mathrm{N}}(t) = -\pi p_{\mathrm{LOS}}\mathrm{D}^2\mathtt{Q}\left(\Upsilon_{\mathrm{N}}(t)\right) + \pi t^{\frac{2}{\alpha_{\mathrm{N}}}} \exp\left(\frac{2\sigma_{\mathrm{N}}^2}{\alpha_{\mathrm{N}}^2} + \frac{2m}{\alpha_{\mathrm{N}}}\right)\left[1 - p_{\mathrm{LOS}}\mathtt{Q}\left(\frac{2\sigma_{\mathrm{N}}^2}{\alpha_{\mathrm{N}}\sigma_{\mathrm{N}}} - \Upsilon_{\mathrm{N}}(t)\right)\right].$$

*Proof.* Proceeds very similarly to Lemma 3 and thus is omitted. □

Note that here $\Lambda_{\mathrm{L}}([0,\infty)) = \lambda_{\mathrm{BS}}\pi p_{\mathrm{LOS}}\mathrm{D}^2$. The probability that there is no point in the interval $[0,\infty)$ is equal to $\exp\left(-\lambda_{\mathrm{BS}}\pi p_{\mathrm{LOS}}\mathrm{D}^2\right)$. This is exactly the probability that there is no point in $\Phi_{\mathrm{L}}$. Let us call the probability that there is at least one point in $\mathcal{N}_{\mathrm{L}}$ to be $\mathrm{B}_{\mathrm{L}}$ .The event that number of points in $\Phi_{\mathrm{N}}$ is zero is empty and thus, $\mathrm{B}_{\mathrm{N}} = 1$.

**Corollary 1.** *Let $\mathcal{N}$ be the point process of propagation losses corresponding to $\Phi_{\mathrm{BS}}$. This point process is a PPP with intensity $\Lambda((0,t]) = \lambda_{\mathrm{BS}}(\mathrm{M}_{\mathrm{L}}(t) + \mathrm{M}_{\mathrm{N}}(t)) = \lambda_{\mathrm{BS}}\mathrm{M}(t)$.*

*Proof.* Follows directly from the Superposition property of PPPs [72, Proposition 1.3.3]. □

**Lemma 5.** *Given that $\mathcal{N}_{\mathrm{L}}$ and $\mathcal{N}_{\mathrm{N}}$ are not empty, the probability density function (PDF) of the distance to the point nearest to origin in these point processes is given by $f_{\mathrm{L}}(t) = \lambda_{\mathrm{BS}}\exp\left(-\lambda_{\mathrm{BS}}\mathrm{M}_{\mathrm{L}}(t)\right)\mathrm{M}_{\mathrm{L}}'(t)/\mathrm{B}_{\mathrm{L}}$ and $f_{\mathrm{N}}(t) = \lambda_{\mathrm{BS}}\exp\left(-\lambda_{\mathrm{BS}}\mathrm{M}_{\mathrm{N}}(t)\right)\mathrm{M}_{\mathrm{N}}'(t)/\mathrm{B}_{\mathrm{N}}$, where $\mathrm{M}_{\mathrm{L}}'(t)$ and $\mathrm{M}_{\mathrm{N}}'(t)$ are given as follows.*

$$\mathrm{M}_{\mathrm{L}}^{'}(t)=\pi p_{\mathrm{LOS}}\Bigg\{\exp\left(\frac{2\sigma_{\mathrm{L}}^{2}}{\alpha_{\mathrm{L}}^{2}}+\frac{2m}{\alpha_{\mathrm{L}}}\right)t^{\frac{2}{\alpha_{\mathrm{L}}}-1}\Bigg[\frac{2}{\alpha_{\mathrm{L}}}\mathtt{Q}\left(\frac{2\sigma_{\mathrm{L}}^{2}}{\alpha_{\mathrm{L}}\sigma_{\mathrm{L}}}-\Upsilon_{\mathrm{L}}(t)\right)$$
$$-\frac{1}{\sqrt{2\pi\sigma_{\mathrm{L}}^{2}}}\exp\left(-\left(\frac{\sqrt{2}\sigma_{\mathrm{L}}}{\alpha_{\mathrm{L}}}-\frac{\Upsilon_{\mathrm{L}}(t)}{\sqrt{2}}\right)^{2}\right)\Bigg]+\frac{\mathrm{D}^{2}}{\sqrt{2\pi}t\sigma_{\mathrm{L}}}\exp\left(-\frac{\Upsilon_{\mathrm{L}}^{2}(t)}{2}\right)\Bigg\},$$

$$\mathrm{M}_{\mathrm{N}}^{'}(t)=\pi p_{\mathrm{LOS}}\Bigg\{\exp\left(\frac{2\sigma_{\mathrm{N}}^{2}}{\alpha_{\mathrm{N}}^{2}}+\frac{2m}{\alpha_{\mathrm{N}}}\right)t^{\frac{2}{\alpha_{\mathrm{N}}}-1}\Bigg[\frac{2}{p_{\mathrm{LOS}}\alpha_{\mathrm{N}}}-\frac{2}{\alpha_{\mathrm{N}}}\mathtt{Q}\left(\frac{2\sigma_{\mathrm{N}}^{2}}{\alpha_{\mathrm{N}}\sigma_{\mathrm{N}}}-\Upsilon_{\mathrm{N}}(t)\right)$$
$$+\frac{1}{\sqrt{2\pi\sigma_{\mathrm{N}}^{2}}}\exp\left(-\left(\frac{\sqrt{2}\sigma_{\mathrm{N}}}{\alpha_{\mathrm{N}}}-\frac{\Upsilon_{\mathrm{N}}(t)}{\sqrt{2}}\right)^{2}\right)\Bigg]-\frac{\mathrm{D}^{2}}{\sqrt{2\pi}t\sigma_{\mathrm{N}}}\exp\left(-\frac{\Upsilon_{\mathrm{N}}^{2}(t)}{2}\right)\Bigg\}.$$

*Proof.* If $l^{*}$ is the point nearest to origin in the point process $\mathcal{N}_{L}$,

$$\begin{aligned}
&\mathbb{P}\left(l^{*}>t\middle|\mathcal{N}_{L}([0,\infty))>0\right)\\
&=\mathbb{P}\left(\mathcal{N}_{L}([0,t))=0\middle|\mathcal{N}_{L}([0,\infty))>0\right)\\
&=\frac{\mathbb{P}\left(\mathcal{N}_{L}([0,t))=0\cap\mathcal{N}_{L}([0,\infty))>0\right)}{\mathbb{P}\left(\mathcal{N}_{L}([0,\infty))>0\right)}\\
&=\mathbb{P}\left(\mathcal{N}_{L}([0,t))=0\cap\mathcal{N}_{L}([t,\infty))>0\right)/\mathrm{B}_{\mathrm{L}}\\
&=\mathbb{P}\left(\mathcal{N}_{L}([0,t))=0\right)\mathbb{P}\left(\mathcal{N}_{L}([t,\infty))>0\right)/\mathrm{B}_{\mathrm{L}}\\
&=\exp\left(-\Lambda_{\mathrm{L}}([0,t)]\right)\left(1-\exp\left(-\Lambda_{\mathrm{L}}([t,\infty)]\right)\right)/\mathrm{B}_{\mathrm{L}}\\
&=\left(\exp\left(-\Lambda_{\mathrm{L}}([0,t)]\right)-\exp\left(-\Lambda_{\mathrm{L}}([0,\infty)]\right)\right)/\mathrm{B}_{\mathrm{L}}.
\end{aligned}$$

Thus, taking the negative derivative of the above expression the PDF $f_{\mathrm{L}}(t)$ is obtained. Similarly, the PDF for the NLOS case can be obtained. □

**Theorem 1.** *The* SNR *coverage of a typical user in the network is given by*

$$\overline{\mathcal{S}}(\tau)\triangleq\mathbb{P}(\mathsf{SNR}_{x,0}>\tau)=\mathbb{E}_{U_{x}}\left[\mathcal{S}(\tau,U_{x})\right],\ \textit{where}\tag{2.11}$$

$$\mathcal{S}(\tau,\mathrm{U}) \approx \sum_{j\in\{\mathrm{L},\mathrm{N}\}} \mathrm{B}_j \zeta(\eta_j,\mathrm{U}) \sum_{n=1}^{\eta_j} (-1)^{n+1} \binom{\eta_j}{n} \times \int_0^{\infty} \exp\left(-\frac{\eta_j \tau n \mathrm{U} l \sigma_n^2}{\mathrm{G}} - \lambda_{\mathrm{BS}} \mathrm{M}_{\bar{j}}(l)\right) f_j(l) \mathrm{d}l,$$

*where* $\mathrm{G} = \mathrm{PN}_{\mathrm{BS}}\mathrm{N}_{\mathrm{UE}}$, $\bar{j} = \mathrm{L}$ *if* $j = \mathrm{N}$ *and vice versa. The terms* $\zeta(.)$, $\mathrm{M}_j(.)$ *and* $f_{(.)}$ *are derived in Proposition 1, Lemma 3, Lemma 4 and Lemma 5.*

*Proof.* See Appendix 2.9.2 □

**Corollary 2.** *Assuming that user density is much larger than BS density, the* SNR *coverage can be approximated by* $\mathcal{S}(\tau, \mathrm{U_M})$.

**Theorem 2.** *In a noise-limited network, the per user rate distribution (or rate coverage) of a* typical *user at origin served by a BS at* $x$ *is given by*

$$\begin{aligned}\mathcal{R}(\tau_r) &\triangleq \mathbb{P}\left(R_{x,0} > \tau_r\right) \\ &= \sum_{n\geq 1} \kappa(n) \mathcal{S}\left(2^{\frac{\tau_r n}{\omega_{\mathrm{MU}} \mathrm{B} \min(n,\mathrm{U_M})}} - 1, \min\left(n, \mathrm{U_M}\right)\right),\end{aligned}$$

*where* $\mathcal{S}(.)$ *was defined in Theorem 1 and* $\kappa(n)$ *is given in (2.8).*

*Proof.* Follows by re-arranging (2.4) and using SNR = SINR. □

Although the above expression is an infinite summation, as verified earlier in [46, 73], it can be accurately represented as a finite summation. For the results in this work, considering the first $\lfloor 12\lambda_{\mathrm{UE}}/\lambda_{\mathrm{BS}} \rfloor$ terms is sufficient. The following definition will be useful when comparing the rate coverage of MU-MIMO with SM and SU-BF.

**Definition 2.** The minimum allowable efficiency of scheme A such that it is guaranteed to outperform scheme B in terms of per user rate for $p$ percentile users (that is users with rate coverage $p$), is given by $\mathcal{O}_{\mathrm{A,B}}(p) = \frac{\mathcal{R}_{\mathrm{B}}^{-1}(p)}{\mathcal{R}_{\mathrm{A}}^{-1}(p)}$, where $\mathcal{R}_{\mathrm{A}}^{-1}$ and $\mathcal{R}_{\mathrm{B}}^{-1}$ are inverse of the rate coverage at $p$ (that is rate thresholds $\tau$ corresponding to $\mathcal{R}(\tau) = p$) for schemes A and B after setting $\omega_{\mathrm{A}} = \omega_{\mathrm{B}} = 1$, where $\omega_{(.)}$ are the efficiency factors for the respective MIMO techniques as defined in (2.4). The per user rate of A cannot stochastically dominate that of B, unless the efficiency of A is at least $\min_p \mathcal{O}_{\mathrm{A,B}}(p)$.

Note that MU-MIMO implementations with different $\mathrm{U_M}$ are considered as separate MIMO schemes in the above definition since they have different efficiency factors.

#### 2.5.2.2 Rate Distribution in an Interference-limited Network

Until now, the analysis focused on noise-limited mmWave cellular networks. In this section, the discussion will be on how to model interference in these networks.

From (2.3), the OCI power at user $u$ served by a BS at $x$ is modelled as

$$I_u = \mathrm{P} \sum_{\substack{y \in \Phi_{\mathrm{BS}} \\ y \neq x}} \sum_{w \in \mathcal{U}_y} \frac{||\overline{\mathbf{h}}_{y,u}^{*} \mathbf{f}_{y,w}^{\mathrm{BB}}||^2}{U_y} = \mathrm{P} \sum_{\substack{y \in \Phi_{\mathrm{BS}} \\ y \neq x}} \sum_{w \in \mathcal{U}_y} \frac{||\mathbf{w}_u^{*} \mathbf{H}_{y,u} \mathbf{F}_y^{\mathrm{RF}} \mathbf{f}_{y,w}^{\mathrm{BB}}||^2}{U_y}.$$

Here, $\mathbf{w}_u = \mathbf{a}_{\mathrm{UE}}(\phi_{x,u})$, $\mathbf{H}_{y,u} = \sqrt{\frac{\mathrm{N_{BS}N_{UE}}}{L(y,u)\eta_{y,u}}} \sum_{i=1}^{\eta_{y,u}} \gamma_{i,y,u} \mathbf{a}_{\mathrm{UE}}(\phi_{i,y,u}) \mathbf{a}_{\mathrm{BS}}^{*}(\theta_{i,y,u})$, $\mathbf{F}_y^{\mathrm{RF}}$ has columns equal to $\mathbf{a}_{\mathrm{BS}}(\theta_{y,w})$ for all $w \in \mathcal{U}_y$, and $\mathbf{f}_{y,w}^{\mathrm{BB}}$ is designed so as

to cancel the multiuser interference of the BS at $y$. All the AOAs and AODs in the above expression are independent of each other. Leveraging the virtual channel approximation for large number of antennas at the BS and UE, interference due to the link between BS at $y$ and user at $w$ on the user $u$ is non-zero if and only if $\phi_{x,u}$ is equal to at least one of the AOA of $\mathbf{H}_{y,u}$ and $\theta_{y,w}$ equals the corresponding AOD. Note that since multiuser interference was cancelled by the ZF precoder, the virtual approximation with an ON/OFF model for inner product of two beam steering vectors gave us a tractable and accurate tool to study SNR distribution in the previous section. However, this model may not be accurate when OCI is incorporated.

The virtual channel approximation quantized the angular space into N sectors, where N is the number of antennas. If two angles lie on either sides of a sector boundary, the inner product of beam steering vectors is zero, which can be a main cause of underestimated interference. In order to alleviate this problem, a notion of side lobe gain, which was also used in [34, 46], is introduced. It is still assumed that the virtual angle space is quantized into N sectors with the angle bisector being a representative of each sector, but the inner product between two beam steering vectors is defined as:

$$\mathbf{a}_{\mathrm{BS}}^{*}(\theta_1)\mathbf{a}_{\mathrm{BS}}(\theta_2) \triangleq \begin{cases} 1 & \text{if } \theta_1 = \theta_2 \\ \rho_{\mathrm{BS}} & \text{otherwise,} \end{cases} \tag{2.12}$$

where $\rho_{\mathrm{BS}} < 1$ introduces a sidelobe gain into the model. Similarly, the inner product for beam steering vectors at UEs with parameter $\rho_{\mathrm{UE}}$ is modeled. Note that setting $\rho_{\mathrm{BS}} = \rho_{\mathrm{UE}} = 0$ reverts back to the virtual channel approximation.

To characterize the interference distribution, the effect of ZF on interfering links and dependence in $p_{\mathrm{MU}}$ and $I_u$ through $\{\mathbf{w}_u\}$ is neglected for tractability. Later it is shown numerically that for a fairly large number of antennas this gives a reasonable approximation. First the following case is dealt: $\eta_{\mathrm{L}} = \eta_{\mathrm{N}} = 1$.

**Proposition 2.** *Assuming the inner product of any two beam steering vectors at BS or UE follow the law given by (2.12), $\eta_{\mathrm{L}} = \eta_{\mathrm{N}} = 1$ and propagation loss on the service link is $l$, the OCI power at the typical user can be modelled as $I_0 = \sum_{y \in \Phi_{\mathrm{BS}}, y \neq x} \mathrm{G}|\gamma_{y,0}|^2 L(y,0)^{-1} \chi_y / U_y$, where $\gamma_{y,0}$ is complex normal random variable with unit variance and zero mean, $U_y$ are i.i.d random variables with distribution given in (2.10) and $\chi_y$ is defined as*

$$\chi_y = \begin{cases} k + (U_y - k)\rho_{\mathrm{BS}}^2 & w.p.\ (\sum_{i=1}^{\mathrm{N_{UE}}} q_{\mathrm{UE},i}^2)\binom{U_y}{k} \sum_{j=1}^{\mathrm{N_{BS}}} q_{\mathrm{BS},j}^{k+1}(1 - q_{\mathrm{BS},j})^{U_y - k} \\ \rho_{\mathrm{UE}}^2(k + (U_y - k)\rho_{\mathrm{BS}}^2) & w.p.\ (1 - \sum_{i=1}^{\mathrm{N_{UE}}} q_{\mathrm{UE},i}^2)\binom{U_y}{k} \sum_{j=1}^{\mathrm{N_{BS}}} q_{\mathrm{BS},j}^{k+1}(1 - q_{\mathrm{BS},j})^{U_y - k}, \end{cases}$$

*for $k = 0, 1, \ldots, U_y$.*

*Proof.* For single path channel, the out-of-cell interference is given by

$$I_0 = \sum_{y \in \Phi_{\mathrm{BS}}, y \neq x} \frac{\mathrm{G}|\gamma_y|^2 L(y,u)^{-1}}{U_y} \sum_{w \in \mathcal{U}_y} ||\mathbf{a}_{\mathrm{UE}}^*(\phi_{x,u}) \mathbf{a}_{\mathrm{UE}}(\phi_{y,u}) \mathbf{a}_{\mathrm{BS}}^*(\theta_{y,u}) \mathbf{a}_{\mathrm{BS}}(\theta_{y,w})||^2.$$

Now using the inner product rule in (2.12) and the fact that all the virtual angles in the above equation are independent and distributed according to Lemma 2, the proposition can be proved. □

**Lemma 6.** *The Laplace functional of the interference power in Proposition 2 conditioned on path loss to the* typical *user at origin from serving BS is* $L(x,u)=l$*, is given by*

$$\begin{aligned}
&L_{I_0,l}(s) \triangleq \mathbb{E}\left[\exp\left(-sI_0\right)|L(x,0)=l\right] \\
&= \exp\left(-\lambda_{\mathrm{BS}}\sum_{n=0}^{\mathrm{U_M}}\tilde{p}(n)\sum_{k=0}^{n}\binom{n}{k}\sum_{i=1}^{\mathrm{N_{BS}}}q_{\mathrm{BS},i}^{k+1}(1-q_{\mathrm{BS},i})^{n-k}\right. \\
&\qquad\left\{\left(\sum_{i=1}^{\mathrm{N_{UE}}}q_{\mathrm{UE},i}^2\right)\int_{t\geq l}\frac{\mathrm{M}^{'}(t)\mathrm{d}t}{1+\frac{tn}{s\mathrm{G}(k+(n-k)\rho_{\mathrm{BS}}^2)}}\right. \\
&\qquad\left.\left.+\left(1-\sum_{i=1}^{\mathrm{N_{UE}}}q_{\mathrm{UE},i}^2\right)\int_{t\geq l}\frac{\mathrm{M}^{'}(t)\mathrm{d}t}{1+\frac{tn}{s\mathrm{G}\rho_{\mathrm{UE}}^2(k+(n-k)\rho_{\mathrm{BS}}^2)}}\right\}\right).
\end{aligned}$$

*where* $\tilde{p}(.)$ *is the distribution of* $U_y$ *for interfering BSs given in (2.10).*

*Proof.* Appendix 2.9.3. □

**Theorem 3.** *The* SINR *coverage of the typical user is given by (2.11) with an extra term* $L_{I_0,l}\left(\frac{\eta_j\tau n\mathrm{U}l}{\mathrm{G}}\right)$ *inside the integral over* $\mathrm{d}l$*.*

*Proof.* Exactly on same lines as Theorem 1. The Laplace functional $L_{I_0,l}(.)$ has been derived in Lemma 6 for single path channel. Upper and lower bounds on $L_{I_0,l}(.)$ for a general number of paths can be found in Appendix 2.9.4. □

From this expression of SINR coverage, the rate coverage can be found similar to Theorem 2. These analytical results will be validated in Section 2.7. In the next section, a brief discussion on the coverage and rate for SM enabled mmWave cellular networks is done. Before that though, a short note on how

to choose $\rho_{\mathrm{UE}}$ and $\rho_{\mathrm{BS}}$ is provided here. Recall that $\mathrm{N_{BS}}\rho_{\mathrm{BS}}^2$ and $\mathrm{N_{UE}}\rho_{\mathrm{UE}}^2$ are the sidelobe gains for beam pattern at BSs and UEs, respectively. An obvious question is whether these parameters depend on the number of antennas and if yes, how should their dependence be modelled?

If $\rho_{(.)}$ were to be a constant, the sidelobe gain will also scale up with an increasing number of antennas. This will violate Lemma 1. Since virtual channel approximation asymptotically tracks physical channel model, $\rho_{\mathrm{BS}}$ and $\rho_{\mathrm{UE}}$ should decrease and eventually vanish with increasing $\mathrm{N_{BS}}$ and $\mathrm{N_{UE}}$, respectively. For a uniform linear array with N antennas, the ratio of the gain of the $i^{\mathrm{th}}$ sidelobe to the main lobe is equal to $|\frac{\sin(0.5\pi(2i+1))}{\mathrm{N}\sin(0.5\pi(2i+1)/\mathrm{N})}|^2$ [76], for $i = 1, 2, \ldots, \lfloor \frac{\mathrm{N}}{2} - 1 \rfloor$. For $i \ll \mathrm{N}$, this ratio is independent of N using the small angle approximation $\sin\theta \approx \theta$. For $i$ on the order of N, this ratio decreases approximately as square of N. The regime in which the ratio is independent of N has about fixed beam width, which corresponds to the beam width in which the small angle approximation of $\sin\theta \approx \theta$ is accurate with $p$ percent relative error. For $p = 1$, $\theta \approx 0.244$ radians. Since the majority of the angular space corresponds to the regime in which the above ratio varies inversely with the square of N, $\rho_{(.)}$ is modeled to linearly decrease with N. The following values are chosen: $\rho_{\mathrm{BS}} = 1/(\sin(0.244)\mathrm{N_{BS}})$ and $\rho_{\mathrm{UE}} = 1/(\sin(0.244)\mathrm{N_{UE}})$. However, in future it is desirable to re-investigate the scaling factor to get a better fit.

## 2.6 Single User Spatial Multiplexing in mmWave Cellular Networks

For spatial multiplexing (SM), every BS transmits more than one stream of data to a single user per resource block. Thus, $\mathrm{N}_s^{\mathrm{UE}} = \mathrm{N}_s^{\mathrm{BS}} = \mathrm{N}_s$, where $\mathrm{N}_s$ is the multiplexing gain. In this section, the focus will be mainly on the multipath diversity approach for SM [27, 55] and not on the polarization approach [55, 56].

### 2.6.1 Spatial Multiplexing: UHF versus mmWave

A brief recap of the theoretically optimal implementation of closed-loop SM in conventional cellular networks is given first, which motivates the main challenges in precoding/combining for SM in mmWave networks. Under unitary power constraint, given the singular value decomposition of the channel matrix $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$, the transmitter pre-multiplies the input symbols with matrix $\mathbf{V}$ and the receiver combines the received signal on all its antennas with matrix $\mathbf{U}^*$, to effectively achieve $\mathrm{N}_s$ parallel channels, where $\mathrm{N}_s$ is the multiplexing gain. Since the channel matrix is either full row rank or full column rank with high probability for sub 6 GHz frequency bands, $\mathrm{N}_s = \min\{\mathrm{N}_{\mathrm{BS}}, \mathrm{N}_{\mathrm{UE}}\}$.

At mmWave frequencies, however, the first challenge is that it is not practically feasible to implement a fully digital precoder and combiner. Using the popular hybrid beamforming approach for mmWave networks [30], the precoder is of the form $\mathbf{F}_x^{\mathrm{RF}}\mathbf{F}_x^{\mathrm{BB}}$, wherein $\mathbf{F}_x^{\mathrm{RF}}$ is generally implemented using

phase shifters and thus has constant magnitude entries. Similarly, there is a constraint for the combiner. Another challenge for implementing SM at mmWave is that the channel is sparse [55, 64] and thus obtaining multiplexing gain on the order of number of antennas is nearly impossible.

A typical implementation of SM using the hybrid beamforming architecture is shown in Fig. 2.1. Assuming perfect channel estimation, and using the system model from Section 2.4, the received signal at user $u$ from BS $x$ is given by

$$\mathbf{y}_u = \mathbf{H}_{x,u}\mathbf{F}_x^{\mathrm{RF}}\mathbf{F}_x^{\mathrm{BB}}\mathbf{s}_u + \mathbf{n} + \mathrm{OCI},$$

where $\mathbf{s}_u$ are transmit symbols of dimension $\mathrm{N}_s \times 1$ with energy per symbol equal to $\mathrm{P}/\mathrm{N}_s$, $\mathbf{n}$ is the noise power (complex Gaussian with zero mean and variance $\sigma_n^2$) and $\mathrm{OCI} = \sum_{y\in\Phi_{\mathrm{BS}}\setminus\{x\}} \mathbf{H}_{y,u}\mathbf{F}_y^{\mathrm{RF}}\mathbf{F}_y^{\mathrm{BB}}\mathbf{s}_y$. We assume equal power allocation to all streams. After RF and baseband combining at the receiver, the processed signal is of the form $\mathbf{W}_u^{\mathrm{BB}^*}\mathbf{W}_u^{\mathrm{RF}^*}\mathbf{y}_u$. When Gaussian symbols are transmitted by all BSs, the achievable rate by treating interference as noise is given as

$$r = \log_2\left|\mathbf{I}_{\mathrm{N_s}} + \frac{\mathrm{P}}{\mathrm{N}_s}\mathbf{R}_{\mathbf{n}}^{-1}\mathbf{H}_{\mathrm{eff}}\mathbf{H}_{\mathrm{eff}}^*\right|,$$

where $\mathbf{H}_{\mathrm{eff}} = \mathbf{W}_u^{\mathrm{BB}^*}\mathbf{W}_u^{\mathrm{RF}^*}\mathbf{H}_{x,u}\mathbf{F}_x^{\mathrm{RF}}\mathbf{F}_x^{\mathrm{BB}}$, $\mathbf{I}_{\mathrm{N}_s}$ is an identity matrix of rank $\mathrm{N}_s$ and

$$\mathbf{R}_{\mathbf{n}} = \sigma_n^2\mathbf{W}_u^*\mathbf{W}_u + \sum_{y\in\Phi_{\mathrm{BS}}\setminus\{x\}} \frac{\mathrm{P}}{\mathrm{N}_{s,y}}\mathbf{H}_{y,\mathrm{eff}}\mathbf{H}_{y,\mathrm{eff}}^*,$$

where $\mathbf{W}_u = \mathbf{W}_u^{\mathrm{RF}}\mathbf{W}_u^{\mathrm{BB}}$ and $\mathbf{H}_{y,\mathrm{eff}} = \mathbf{W}_u^*\mathbf{H}_{y,u}\mathbf{F}_y^{\mathrm{RF}}\mathbf{F}_y^{\mathrm{BB}}$. The precoding-combining algorithm proposed in [27] is used for Monte Carlo simulations. Assuming that

an equal fraction of resource is allocated to each UE connected to a BS, the per user rate is defined as $R_{x,u} = \omega_{\mathrm{SM}} \mathrm{B}r/N_x$, where $\omega_{\mathrm{SM}}$ is the efficiency factor for SM and recalling that $N_x$ is the total number of users associated with the BS at $x$. Note that similar to the MU case, $\omega_{\mathrm{SM}}$ is dependent on several network parameters like number of antennas, the channel parameters, number of streams, etc. but this is dropped in the notation for convenience. Sum rate is defined as the total bits per second transmitted by a BS in Section 2.5. Based on this definition, the sum rate for the SM enabled mmWave network is defined as $R_x = \omega_{\mathrm{SM}} \mathrm{B}r$.

### 2.6.2 Heuristic Comparison of Coverage and Rate for MU-MIMO and SM

In this section, the SNR for spatial multiplexing and MU-MIMO is denoted with a superscript SM and MU. Round robin scheduling and $\omega_{\mathrm{MU}} = \omega_{\mathrm{SM}} = 1$ will be assumed in this section. From Lemma 1, for a large number of antennas the singular values of $\mathbf{H}_{\mathrm{eff}}\mathbf{H}_{\mathrm{eff}}^{*}$ converge to $\frac{\mathrm{N_{BS}N_{UE}}|\gamma_{i,x,u}|^2}{L(x,u)\eta_{x,u}}$. Thus, the ratio $\frac{\mathsf{SNR}_{i,x,u}^{\mathrm{SM}}}{\mathrm{G}} \xrightarrow{d} \frac{|\gamma_{i,x,u}|^2 L(x,u)^{-1}}{\eta_{x,u}\mathrm{N}_s}$. From Remark 3, the ratio $\frac{\mathsf{SNR}_{x,u}^{\mathrm{MU}}}{\mathrm{G}} \xrightarrow{d} \frac{|\gamma_{i_m,x,u}|^2 L(x,u)^{-1}}{\eta_{x,u}\mathrm{N}_s}$, where $i_m = \arg\max_i \gamma_{i,x,u}$. Since $\gamma_{i,x,u} \overset{\mathrm{st}}{\leq} \max_i \gamma_{i,x,u}$, one can conclude that in the limit as $\mathrm{N_{BS}} \to \infty$ and $\mathrm{N_{UE}} \to \infty$, $\frac{\mathsf{SNR}_{i,x,u}^{\mathrm{SM}}}{\mathrm{G}} \overset{\mathrm{st}}{\leq} \frac{\mathsf{SNR}_{x,u}^{\mathrm{MU}}}{\mathrm{G}}$ for all $i \in \{1, \ldots, \mathrm{N}_s\}$.

The above discussion hints that for many antennas at BS and UE, the SNR with MU-MIMO stochastically dominates the SNR on each stream of SM. If the network were to be noise-limited, the per user and sum rates with MU-

MIMO will be higher than SM for a large number of antennas and the same number of streams. Now, let us consider how this result might be affected by OCI. As the number of antennas become large, the effect of zero forcing on the interfering streams is negligible for both MU-MIMO and SM (since the virtual channel approximation in [54] starts to more closely model the actual channel). Thus, if the number of streams transmitted by the BS with SM and MU-MIMO are the same, the interference statistics with MU-MIMO and SM would be similar and one would expect that MU-MIMO still outperforms SM for a large number of antennas at BSs and UEs.

For a finite number of antennas the ZF penalty may be non-negligible. It is expected that the ZF penalty with SM will be less than MU-MIMO since there are more sidelobes that need to be suppressed with MU-MIMO. Thus, the above SNR dominance result holds given that the number of antenna is large enough such that the effect of the smaller ZF penalty with SM does not reverse the inequalties. For a finite number of antennas, it is neither obvious nor analytically tractable to conjecture as to whether the per user and sum rate of SM would dominate or whether MU-MIMO would. We, thus, rely on Monte Carlo simulations for SM while comparing with our validated analytical model for MU-MIMO and SU-BF.

## 2.7 Numerical Results

In this section, we first validate the SNR, SINR and rate coverage analysis from Section III. Next, we compare the per user and sum rate for SU-BF,

Table 2.1: Simulation parameters

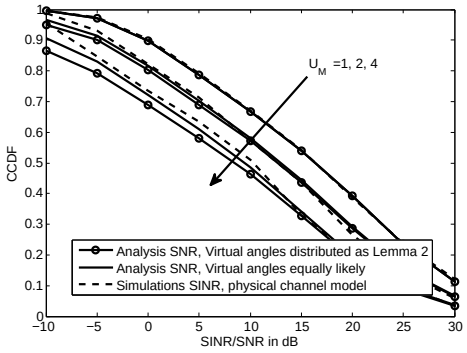
| **Parameter** | **Value(s)** | **Parameter** | **Value(s)** |
|---|---|---|---|
| $f_c$ | 73 GHz [56] | B | 1 GHz [56] |
| $p_{\text{LOS}}$, D | 0.11, 200 m [46] | $\sigma_n^2$ | $-174 + 10\log_{10}\text{B} + 10$ dBm |
| $\alpha$ (LOS, NLOS) | 2, 3.3 [56] | $\xi$ (LOS, NLOS) | 5.2, 7.6 [56] |
| $\lambda_{\text{UE}}$ | $500/\text{km}^2$ | $\lambda_{\text{BS}}$ | $60/\text{km}^2$ |
| $\text{N}_{\text{UE}}$ | 16 [24, 77] | $\text{N}_{\text{BS}}$ | 64 [24, 77] |
| P | 30 dBm [17] | $\eta_{\text{L}}$, $\eta_{\text{N}}$ | 1,3 [24, 52, 65] |

MU-MIMO and SM with fixed number of BSs per unit area as well as fixed power consumption per unit area. The default parameters used for generating the results are given in Table 2.1. The efficiency factors $\omega_{\text{MU}}$ and $\omega_{\text{SM}}$ are implementation specific and estimating these is not the focus of this study. Thus, we set the efficiency parameters to 1 and use Definition 2 for quantifying the allowable relative efficiency.

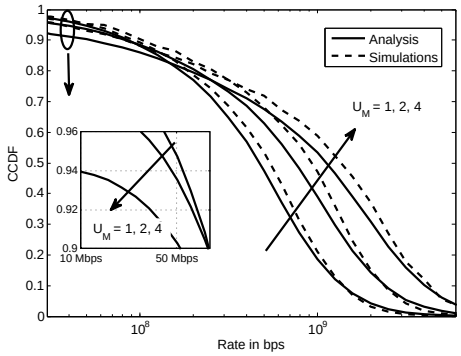
### 2.7.1 Coverage and Rate with MU-MIMO: Validation and Trends

#### 2.7.1.1 Cases Where Interference is Negligible

Fig. 2.2(a) shows the validation of the SNR coverage formula in Theorem 1. As can be seen from the figure, the analysis is a tight approximation with the simulations using the physical channel model even when the virtual angles are equally likely, in which case we have much simplified analytical expressions as compared to when the distribution is as given in Lemma 2. Henceforth, all analysis plots will be with equally likely virtual angles. As expected, the match loosens as $\text{U}_{\text{M}}$ approaches $\text{N}_{\text{BS}}$ and $\text{N}_{\text{UE}}$. With increasing




(a) Comparison of SNR analysis with SINR simulation.

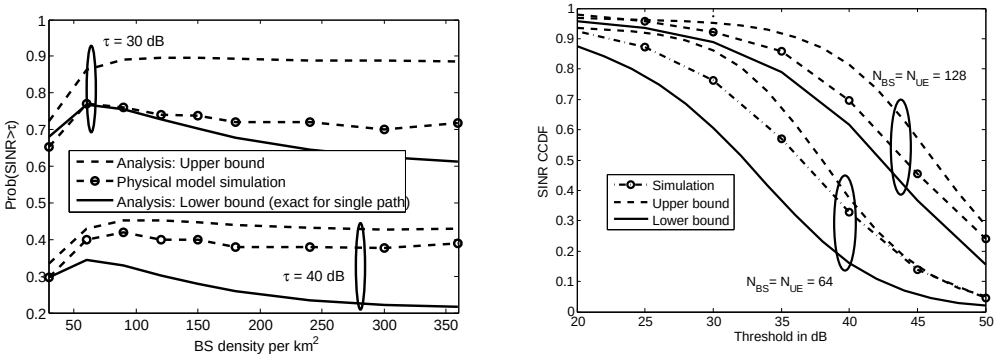



(b) Comparison of per user rate

Figure 2.2: Validation of SNR analysis in noise-limited scenario shows a tight match with the physical channel model simulations. Tradeoff between SINR and rate coverage is also shown with MU-MIMO.

$\mathrm{U_M}$, the SINR coverage decreases since the transmit power is split amongst the multiple users served by the BS. However, as seen from Fig. 2.2(b), the median and peak per user rate increases with MU-MIMO. This is due to the fact that in round robin scheduling, each user connected to BS at $x$ now gets $\min(\mathrm{U_M}, N_x)$ times more slots to transmit. A re-interpretation of the above result can be made in terms of minimum allowable efficiencies. For example, $\mathcal{O}_{\{\mathrm{U_M}=2\},\{\mathrm{U_M}=1\}}(0.5) = 62.67\%$ and $\mathcal{O}_{\{\mathrm{U_M}=4\},\{\mathrm{U_M}=1\}}(0.5) = 42.73\%$. This means that if the efficiency of implementing MU-MIMO with $\mathrm{U_M} = 2$ is at least 62.67% of the efficiency with $\mathrm{U_M} = 1$, then it is beneficial to employ MU-MIMO with $\mathrm{U_M} = 2$ over SU-BF in terms of the median rates.

Since SINR decreases with $\mathrm{U_M}$, the trend for cell edge rates is exactly opposite to peak and median rates. Note that in [59], it was shown that cell




(a) Validation of SINR, $\eta_{\mathrm{L}} = \eta_{\mathrm{N}} = 1$, $\mathrm{N_{BS}} = \mathrm{N_{UE}} = 64$.

(b) Validation of $\rho_{(.)}$ and bounds on Laplace functional for $\eta_{\mathrm{L}} = 2$, $\eta_{\mathrm{N}} = 3$.

Figure 2.3: Validation in interference limited setting for $\mathrm{U_M} = 4$ shows that the upper and lower bounds are within $\pm 5$ dB of the actual simulations. SINR coverage has a non-monotonic trend with BS density.

edge rates can improve with MU-MIMO. However, the main difference in their model is the user selection and scheduling. In [59], there is a high priority user scheduled in a time slot and additional users are served using MU-MIMO only if the expected sum proportional fair metric does not increase due to addition of more users. This protects the rates achieved by cell edge users. The result in Fig. 2.2(b), thus, highlights the importance of user selection and scheduling to protect the rates achieved by cell edge users with multiuser transmission.

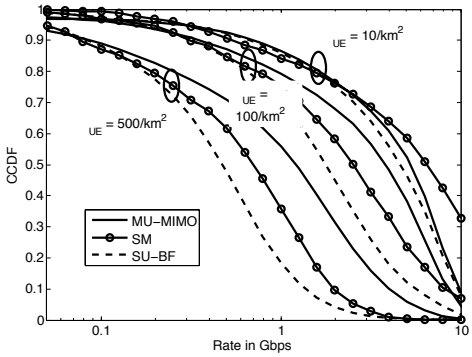
#### 2.7.1.2 Cases Where Interference is Not Negligible

Fig. 2.3(a) shows the validation of SINR coverage formula in Theorem 3 for single path scenario. In order to present a case where interference effects are not negligible we consider a network at 28 GHz band with 200 MHz band-

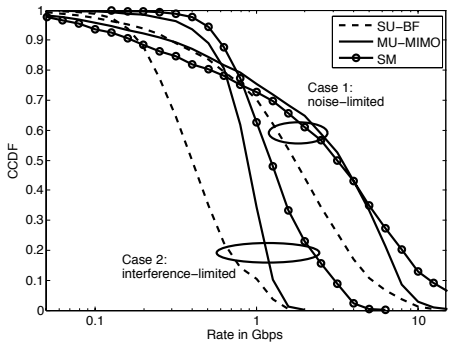
width, a less blocked scenario with $p_{\text{LOS}} = 0.5$ for D = 200m and much higher $\lambda_{\text{UE}} = 1000/\text{km}^2$. As per discussion in Section 2.5, $\rho_{\text{BS}} = 1/(\text{N}_{\text{BS}} \sin(0.244))$ and $\rho_{\text{UE}} = 1/(\text{N}_{\text{UE}} \sin(0.244))$. Fig. 2.3(a) shows that increasing BS density does not necessarily improve coverage. This trend is similar to that observed in [34] and shows the presence of an optimal BS density in terms of SINR coverage. Approximate analytical results in Lemma 6 and Appendix 2.9.4 capture the essential non-monotonic trend shown with the simulations. Fig. 2.3(b) further validates the analysis in Appendix 2.9.4 for multipath scenario as well as shows a decreasing gap with the physical model simulations as the number of antennas grows large. Both these plots build confidence in the analysis and derived insights. Using analysis, it can be found that optimum BS density for $\text{U}_{\text{M}} = 1, 2$ and 4 decreases as $82, 72$ and 63 BSs/km$^2$. Thus, with increasing $\text{U}_{\text{M}}$ the optimum BS density reduces due to increasing interference in the network.

### 2.7.2 Comparing Per User and Sum Rate for SU-BF, MU-MIMO and SM

The gains with SM and MU-MIMO are fundamentally driven by distinct network parameters. For example, having more number of multipaths (or larger $\eta_{\text{L}}$ and $\eta_{\text{N}}$) increases the rank of the channel and thus enables transmitting more number of streams with single user SM, given that there are enough RF chains at the transmitter and the receiver. However, this does not necessarily help in having more multi-user streams. On the other hand, having low load reduces the possible gain with MU-MIMO even if each BS is equipped




(a) Sparse multipath: MU-MIMO outperforms except for very low loads. $\eta_{\mathrm{L}} = 2$, $\eta_{\mathrm{N}} = 3$




(b) Enabling 4 stream SM for $\lambda_{\mathrm{UE}} = 100/\mathrm{km}^2$ shows benefit over MU-MIMO for median/peak rates. $\eta_{\mathrm{L}} = 10$, $\eta_{\mathrm{N}} = 12$, ,

Figure 2.4: Comparison of MIMO techniques with fixed BS density. $\mathrm{U_M} = 4$

with a large number of RF chains due to the fact that there are not many users to schedule simultaneously per BS. This does not however affect SM in terms of the number of streams per user. Thus, sufficiently low load and high multipaths may cause SM to outperform MU-MIMO given that there are enough RF chains at the BSs and UEs. This can be seen in Figures 2.4(a) and 2.4(b). The plots for MU-MIMO and SU-BF in Figure 2.4(a) are with analysis. The plots for SM in Figure 2.4(a) and the entire Figure 2.4(b) is using Monte-Carlo simulations. Note that our analytical model is valid for $\eta_{\mathrm{L}}, \eta_{\mathrm{N}} \gg \mathrm{N_{UE}}$ and not for $\eta_{\mathrm{L}}, \eta_{\mathrm{N}}$ close to $\mathrm{N_{UE}}$, which is the case in Figure 2.4(b).

Figure 2.4(a) shows that for moderate and low user densities (which corresponds to $\lambda_{\mathrm{UE}} = 500/\mathrm{km}^2$ and $\lambda_{\mathrm{UE}} = 100/\mathrm{km}^2$) MU-MIMO outperforms SM and SU-BF. However, for very low load (corresponds to 10 UEs/km$^2$) SM outperforms MU-MIMO. This result is due to the fact that although SM

can offer 2 streams per user but MU-MIMO cannot provide gains since per $km^2$ there are only 10 users that can associate with 60 possible BSs and the probability that a BS connects to more than 1 user is very low. Since our analytical model slightly loose estimates for low SNR users, we compare the cell edge rates using simulations only. The cell edge rates are quite close for the three schemes, although SM and SU-BF slightly outperform MU-MIMO. For low loads, SM is slightly better than SU-BF in terms of cell edge rates. Considering that overhead with MU-MIMO could be the highest, this trend will be more exaggerated afte considering these factors. A better scheduling will be indeed important for protecting cell edge rates with MU-MIMO.

Figure 2.4(b) shows the impact of high multipath on the comparison insights. As was observed in Figure 2.4(a), MU-MIMO outperformed SM for $\lambda_{\mathrm{UE}} = 100/\mathrm{km}^2$ when multipath was low. For the same network parameters, that lead to a noise-limited case, increasing the multipath to $\eta_{\mathrm{L}} = 10$ and $\eta_{\mathrm{N}} = 12$ gives higher rates with SM for even 30 percentile users. This is again due to the fact that since there are 4 RF chains at UEs and BSs, SM can support 4 streams per user. However, since there are about 1.7 UEs per BS, BSs can only transmit to about 2 UEs per time slot on an average with MU-MIMO. Further the increased multipath leads to higher ZF penalty for MU-MIMO. Similar trend is observed in the interference-limited scenario ($f_c = 28$ GHz, B $= 100$ MHz, $p_{\mathrm{LOS}} = 0.5$). Since a low blockage scenario is considered, the 4 streams per UE are LOS links with very high probability. Thus, the gains with SM look slightly exaggerated in the interference-limited

case. Also note that having a large multipath as considered here could be a unlikely scenario in outdoor mmWave networks [24] but it is interesting to consider from an analytical perspective.

A re-interpretation of the above plots can be made in terms of minimum allowable efficiency of MU-MIMO to outperform SM or SU-BF. For example, when $\lambda_{\mathrm{UE}} = 500/\mathrm{km}^2$ in Figure 2.4(a), MU-MIMO outperforms SM in terms of median users if its efficiency factor is more than 58%. Similarly, such numbers can be extracted for other plots using Definition 2. As mentioned earlier, a separate study on estimating these efficiency factors is needed to make a strong claim on comparison of these MIMO techniques.

The above comparison results were for fixed BS density and the same number of antennas across different schemes. However, with an increasing number of RF chains, the power consumed per BS also increases. In the hybrid precoding as shown in Fig. 2.1, each RF chain is connected to all antennas through phase shifters. Thus, with increasing number of RF chains the number of phase shifters grows proportionally with the number of antennas, and effectively the power consumption is also increased. Let $\nu(\mathrm{N_{RF}})$ denote the ratio of power consumed at a BS with $\mathrm{N_{RF}}$ RF chains to a BS with 1 RF chain. A ballpark value of $\nu$ can be found to be 1.38 for $\mathrm{N_{BS}} = 64$ and $\mathrm{N_{RF}} = \mathrm{U_M} = 2$ based on the power consumption model in [69] (refer [78] for a discussion on this).

We now scale up the BS density of SU-BF by exactly a factor of $\nu$. Note that UEs need to use only single RF chain for SU-BF and MU-MIMO

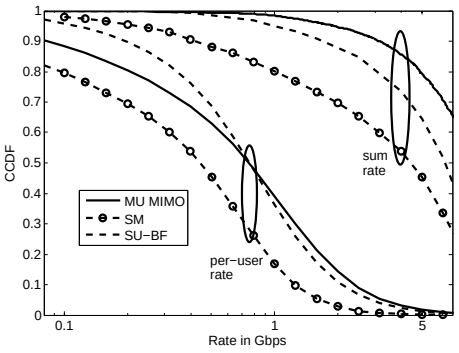



Figure 2.5: Comparison of per user and sum rate with fixed power consumption. A denser SU-BF network outperforms MU-MIMO and SM in terms of cell edge per user rates but MU-MIMO performs the best in terms of sum rates.

with hybrid precoding. However, UEs need multiple RF chains for SM with hybrid precoding architecture. Thus, for fair comparison considering power consumption model in [69] we reduce the $N_{UE}$ to 7 for SM. As can be seen from Fig. 2.5, the gain in per user data rates with MU-MIMO and SM diminishes or completely vanishes if the SU-BF network has 1.38 times denser deployment on an average. Fig. 2.5 shows that MU-MIMO still has significantly higher sum rates than for a denser SU-BF network. However, per user cell edge rates with a denser SU-BF network are higher in this case. To quantify the cell edge gains in per user rates $\mathcal{O}_{\mathrm{MU,SU}}(0.95) = 315\%$, which is huge and strengthens our conclusion that a denser SU-BF network outperforms MU-MIMO in terms of cell edge rates. Also note that $\mathcal{O}_{\mathrm{MU,SU}}(0.5) = 99\%$, which implies that most likely even the median gains with SU-BF will be better after incorporating the channel acquisition overheads. However, in terms of

sum rates $\mathcal{O}_{\mathrm{MU,SU}}(0.5) = 73\%$ which implies that median rates with MU-MIMO can still be higher as long as the efficiency is more than 73% of SU-BF efficiency.

## 2.8 Summary

In this chapter, the coverage and rate performance of hybrid beamforming enabled multi-user (MU) MIMO was compared with single-user spatial multiplexing (SM) and single-user analog beamforming (SU-BF). A stochastic geometry model for coverage and rate analysis was proposed for MU-MIMO mmWave cellular networks, taking into account important mmWave-specific hardware constraints for hybrid analog/digital precoders and combiners, and a blockage-dependent channel model which is sparse in angular domain. The analytical results highlight the coverage, rate and power consumption trade-offs in multiuser mmWave networks. With perfect channel state information at the transmitter and round robin scheduling, MU-MIMO is usually a better choice than SM or SU-BF in mmWave cellular networks. This observation, however, neglected any overhead due to channel acquisition or computational complexity. Incorporating the impact of such overheads, our results are re-interpreted so as to quantify the minimum allowable efficiency of MU-MIMO to provide higher rates than SM or SU-BF.

The analytical model in this work demonstrates the utility of the virtual channel approximation to incorporate different precoder and combiner constraints in network level analysis of dense MIMO cellular networks with

many antennas. It would be beneficial to get tighter bounds on the Laplace functional of the out-of-cell interference. The analytical model can also be extended to incorporate more realistic cross-polarized uniform planar arrays instead of ULA. Another important issue that needs to be addressed is to incorporate the effects of imperfect channel state information in the analytical model. Since MU-MIMO requires more channel state information at the transmitter, imperfect channel knowledge may affect the performance of MU-MIMO more than SM or SU-BF. It is essential to know whether this would overshadow the benefits of MU-MIMO over SM and SU-BF observed in this chapter.

## 2.9 Appendices

### 2.9.1 Derivation of Zero Forcing Penalty in Proposition 1

For simplicity in notation, let us denote by $\theta_j^i$ and $\phi_j^i$ as the AOD and AOA on the $j^{\text{th}}$ path from/to the BS at $x$ under consideration to/from the $i^{\text{th}}$ user, $i \in \{1, \ldots, \mathrm{U}\}$, served by the BS, respectively. $\overline{\overline{\mathbf{H}}}_{x,u}$ is equal to (2.7) when all of the following events are true.

- $E_1$ : $\mathbf{a}^*_{\mathrm{UE}}(\phi_1^k)\mathbf{a}_{\mathrm{UE}}(\phi_j^k)\mathbf{a}^*_{\mathrm{BS}}(\theta_j^k)\mathbf{a}_{\mathrm{BS}}(\theta_1^1) = 0$ for all $j \in \{1, \ldots, \eta_k\}$ and $k \in \{2, \ldots, \mathrm{U}\}$.
- $E_2$ : $\mathbf{a}^*_{\mathrm{UE}}(\phi_1^1)\mathbf{a}_{\mathrm{UE}}(\phi_j^1)\mathbf{a}^*_{\mathrm{BS}}(\theta_j^1)\mathbf{a}_{\mathrm{BS}}(\theta_1^k) = 0$ for all $j \in \{1, \ldots, \eta_1\}$ and $k \in \{2, \ldots, \mathrm{U}\}$.
- $E_3$ : $\mathbf{a}^*_{\mathrm{UE}}(\phi_1^1)\mathbf{a}_{\mathrm{UE}}(\phi_j^1)\mathbf{a}^*_{\mathrm{BS}}(\theta_j^1)\mathbf{a}_{\mathrm{BS}}(\theta_1^1) = 0$ for all $j \in \{2, \ldots, \eta_1\}$.

Note that probability of $p_{\mathrm{ZF}} = 1$ is given by $\mathbb{P}(E_1 \cap E_2 \cap E_3)$. Using the ON/OFF nature of inner products of beamsteering vectors with virtual channel approximation, we can re-write the above conditions as

- $\mathcal{E}_1 = A_1 \cap A_2$, where $A_1 = \bigcap_{k=2}^{\mathrm{U}} \{\theta_1^1 \neq \theta_1^k\}$ and $A_2 = \bigcap_{k=2}^{\mathrm{U}} \bigcap_{j=2}^{\eta_k} \{\phi_1^k \neq \phi_j^k\} \cup \{\theta_j^k \neq \theta_1^1\}$.
- $\mathcal{E}_2 = A_1 \cap A_3$, where $A_3 = \bigcap_{k=1}^{\mathrm{U}} \bigcap_{j=2}^{\eta_1} \{\phi_1^1 \neq \phi_j^1\} \cup \{\theta_j^1 \neq \theta_1^k\}$.

Note that $\mathbb{P}(E_1 \cap E_2 \cap E_3) = \mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2) = \mathbb{P}(A_1 \cap A_2 \cap A_3)$. Conditioning on $\theta_1^1$, $A_2$ is independent of $A_1$ and $A_3$. Using (a) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$, (b) all distinct AOA or AOD are independently distributed as per the distribution given in Lemma 2, (c) $\bigcap_{k=1}^{\mathrm{U}} \bigcap_{j=2}^{\eta_1} \{\phi_1^1 \neq \phi_j^1\} \cup \bigcap_{k=1}^{\mathrm{U}} \bigcap_{j=2}^{\eta_1} \{\theta_j^1 \neq \theta_1^k\} \subset \bigcap_{k=1}^{\mathrm{U}} \bigcap_{j=2}^{\eta_1} \{\phi_1^1 \neq \phi_j^1\} \cup \{\theta_j^1 \neq \theta_1^k\}$ and (d) for a highly dense network, the probability that the BS is serving a LOS UE is expected to be close to $p_{\mathrm{LOS}}$ since the association region of a BS is almost surely covered by the ball of radius D centered at the BS, the required lower bound on the probability of $p_{\mathrm{ZF}} = 1$ is derived, also given by $\zeta(.)$. In order to get the more simplified expression in Remark 1, the term $\mathrm{D}_j(.)$ in Proposition 1 needs to be simplified. For equally likely virtual angles, this can be found using the following Lemma, which we propose.

**Lemma 7.** *Pick U numbers that take values in range* $\{1, \ldots, N\}$. *Repetition of values is allowed and order is important. The probability that the first* $U_1$ *numbers are mutually exclusive from the remaining* $U_2 = U - U_1$ *is given by*

$\mathcal{P}$, *where*

$$\mathcal{P} = \sum_{d=1}^{U_1} \binom{N}{d} \frac{(N-d)^{U_2}}{N^{U_1+U_2}} \sum_{i=0}^{d} \binom{d}{i} (-1)^i (d-i)^{U_1}.$$

The idea is to condition that there are $d$ distinct values in first $U_1$ numbers, in which case the remaining $U_2$ numbers can take values in $(N-d)^{U_2}$ ways. Further the number of ways in which first $U_1$ numbers take $d$ distinct values can be found using inclusion exclusion principle, which is given by the inner summation.

### 2.9.2 Proof of Theorem 1

Let $l_{\mathrm{L}}^*$ and $l_{\mathrm{N}}^*$ denote the points closest to origin in $\mathcal{N}_{\mathrm{L}}$ and $\mathcal{N}_{\mathrm{N}}$, respectively. Using Lemma 5, the probability of associating with a LOS BS is given by

$$\begin{aligned} \mathrm{A_L} &= \mathrm{B_L} \int_0^\infty \mathbb{P}\left(l_{\mathrm{N}}^* > t\right) f_{\mathrm{L}}(t) \mathrm{d}t \\ &= \mathrm{B_L} \int_0^\infty \exp\left(-\lambda_{\mathrm{BS}} \mathrm{M_N}(t)\right) f_{\mathrm{L}}(t) \mathrm{d}t. \end{aligned}$$

Similarly, the probability of associating with NLOS BS is given by $\mathrm{A_N} = \mathrm{B_N} \int_0^\infty \exp\left(-\lambda_{\mathrm{BS}} \mathrm{M_L}(t)\right) f_{\mathrm{N}}(t) \mathrm{d}t$. Similar to Lemma 3 in [34], the PDF of propagation loss to associated BS given that the association is of type LOS, is given by $\tilde{f}_{\mathrm{L}}(t) = \frac{\mathrm{B_L}}{\mathrm{A_L}} f_{\mathrm{L}}(t) \exp\left(-\lambda_{\mathrm{BS}} \mathrm{M_N}(t)\right)$. Similarly, the PDF of propagation loss given the associated BS is NLOS is given by $\tilde{f}_{\mathrm{N}}(t) = \frac{\mathrm{B_N}}{\mathrm{A_N}} f_{\mathrm{N}}(t) \exp\left(-\lambda_{\mathrm{BS}} \mathrm{M_L}(t)\right)$. Define $\mathcal{S}(\tau, \mathrm{U}) \triangleq \mathbb{P}\left(\mathsf{SNR}_{x,0} > \tau | U_x = \mathrm{U}\right)$. Thus, $\overline{\mathcal{S}}(\tau) = \mathbb{E}_{U_x = \mathrm{U}}\left[\mathcal{S}(\tau, \mathrm{U})\right]$. By

the law of total probability,

$$\begin{aligned}
\mathcal{S}(\tau, \mathrm{U}) &= \mathrm{A_L}\mathbb{P}\left(\mathsf{SNR}_{x,0} > \tau \middle| \text{LOS connection}\right) \\
&\quad + \mathrm{A_N}\mathbb{P}\left(\mathsf{SNR}_{x,0} > \tau \middle| \text{NLOS connection}\right) \\
&\overset{(a)}{\approx} \mathrm{A_L}\int_0^\infty \mathbb{P}\left(|\gamma_{i_m,x,0}|^2 > \frac{\eta_{\mathrm{L}}\tau \mathrm{U}t\sigma_n^2}{p_{\mathrm{MU}}\mathrm{G}}\right)\tilde{f}_{\mathrm{L}}(t)\mathrm{d}t \\
&\quad + \mathrm{A_N}\int_0^\infty \mathbb{P}\left(|\gamma_{i_m,x,0}|^2 > \frac{\eta_{\mathrm{N}}\tau \mathrm{U}t\sigma_n^2}{p_{\mathrm{MU}}\mathrm{G}}\right)\tilde{f}_{\mathrm{N}}(t)\mathrm{d}t \\
&= \mathrm{B_L}\int_0^\infty \mathbb{P}\left(|\gamma_{i_m,x,0}|^2 > \frac{\eta_{\mathrm{L}}\tau \mathrm{U}t\sigma_n^2}{p_{\mathrm{MU}}\mathrm{G}}\right)e^{-\lambda_{\mathrm{BS}}\mathrm{M_N}(t)}f_{\mathrm{L}}(t)\mathrm{d}t \\
&+ \mathrm{B_N}\int_0^\infty \mathbb{P}\left(|\gamma_{i_m,x,0}|^2 > \frac{\eta_{\mathrm{N}}\tau \mathrm{U}t\sigma_n^2}{p_{\mathrm{MU}}\mathrm{G}}\right)e^{-\lambda_{\mathrm{BS}}\mathrm{M_L}(t)}f_{\mathrm{N}}(t)\mathrm{d}t,
\end{aligned}$$

where (a) is obtained using Proposition 1. Note that the first integral is the probability that $\mathsf{SNR}$ exceeds the threshold and there is LOS connection, whereas the second term is for NLOS connection. Let us consider the probabilities in each of these two terms separately.

$$\begin{aligned}
&\mathbb{P}\left(|\gamma_{i_m,x,0}|^2 > \frac{\eta_{\mathrm{L}}\tau \mathrm{U}t\sigma_n^2}{p_{\mathrm{MU}}\mathrm{G}}\right) \\
&= \mathbb{P}\left(p_{\mathrm{MU}} = 1\right)\mathbb{P}\left(|\gamma_{i_m,x,0}|^2 > \frac{\eta_{\mathrm{L}}\tau \mathrm{U}t\sigma_n^2}{p_{\mathrm{MU}}\mathrm{G}}\middle| p_{\mathrm{MU}} = 1\right) \\
&\overset{(b)}{=} \zeta(\eta_{\mathrm{L}}, \mathrm{U})\mathbb{P}\left(|\gamma_{i_m,x,0}|^2 > \frac{\eta_{\mathrm{L}}\tau \mathrm{U}t\sigma_n^2}{\mathrm{G}}\right),
\end{aligned}$$

where (b) is obtained from distribution of $p_{\mathrm{MU}}$ in Proposition 1.

Further, using the distribution of maximum of $\eta_{\mathrm{L}}$ exponential random variables for $|\gamma_{i_m,x,0}|^2$,

$$\mathbb{P}\left(|\gamma_{i_m,x,0}|^2 > \frac{\eta_{\mathrm{L}}\tau \mathrm{U}t\sigma_n^2}{\mathrm{G}}\right)$$

$$= \sum_{n=1}^{\eta_{\mathrm{L}}} (-1)^{n+1} \binom{\eta_{\mathrm{L}}}{n} \exp\left(-\eta_{\mathrm{L}} \tau n \mathrm{U} t \sigma_n^2 / \mathrm{G}\right).$$

Similarly, we can find the NLOS probability term, which completes the proof.

### 2.9.3 Proof of Lemma 6

The Laplace functional of the out-of-cell interference to a user at origin, given the path loss to the serving BS, is defined as $L_{I_0,l}(s) \triangleq \mathbb{E}\left[\exp\left(-sI_0\right) | L(x,0) = l\right]$.

$$\begin{aligned}
L_{I_0,l}(s) &= \mathbb{E}\left[\exp\left(-s \sum_{y \in \Phi_{\mathrm{BS}}, y \neq x} \frac{\mathrm{G}|\gamma_{y,0}|^2 \chi_y}{L(y,0) U_y}\right) \middle| L(x,0) = l\right] \\
&\overset{(a)}{=} \mathbb{E}\left[\exp\left(-s \sum_{t \in \mathcal{N}, t \geq l} \frac{\mathrm{G}|\gamma_t|^2 t^{-1} \chi_t}{U_t}\right)\right] \\
&\overset{(b)}{=} \mathbb{E}\left[\prod_{t \in \mathcal{N}, t \geq l} \exp\left(-\frac{s\mathrm{G}|\gamma_t|^2 t^{-1} \chi_t}{U_t}\right)\right] \\
&= \mathbb{E}\left[\prod_{t \in \mathcal{N}, t \geq l} \mathbb{E}_{|\gamma_t|^2}\left[\exp\left(-\frac{s\mathrm{G}|\gamma_t|^2 t^{-1} \chi_t}{U_t}\right)\right]\right] \\
&\overset{(c)}{=} \mathbb{E}\left[\prod_{t \in \mathcal{N}, t \geq l} \frac{1}{1+\psi_t}\right] \\
&\overset{(d)}{=} \exp\left(-\int_l^\infty \left(1 - \mathbb{E}_{\psi_t}\left[\frac{1}{1+\psi_t}\right]\right) \Lambda(\mathrm{d}t)\right) \\
&= \exp\left(-\int_l^\infty \left(\mathbb{E}_{\psi_t}\left[\frac{1}{1+\psi_t^{-1}}\right]\right) \Lambda(\mathrm{d}t)\right),
\end{aligned}$$

where (a) is obtained by displacing each point $y \in \Phi_{\mathrm{BS}}, y \neq x$ to $L(y,0) = t \in \mathcal{N}$, $t \geq l$. Note that $\gamma_{y,0}, U_y$ and $\chi_y$ are independent marks of $y \in \Phi_{\mathrm{BS}}$, whose distributions are themselves independent of the location $y$. After one

to one mapping of each point $y \in \mathbb{R}^2$ to $t \in \mathbb{R}^+$ and each mark to itself, we associate each feasible point $t \in \mathcal{N}$ with independent marks $\gamma_t, U_t$ and $\chi_t$, with same distribution as the corresponding earlier marks. Here, (b) is obtained using independence of the marks of the displaced PPP and (c) since $\gamma_t$ are exponentially distributed random variables with unit mean and $\psi_t = \frac{s\mathrm{G}t^{-1}\chi_t}{U_t}$. Using the PGFL (probability generating functional) [72] we obtain (d). Using the distribution of $U_y$ and $\chi_t$, we get the required result.

### 2.9.4 Laplace Functional of Out-of-cell Interference for General Number of Paths

The out-of-cell interference from a BS at $y$ to user at origin, served by BS at $x$ is given by

$$I_{y,0} = \frac{\mathrm{G}L(y,0)^{-1}}{\eta_{y,u}U_y} \sum_{w \in \mathcal{U}_y} || \sum_{j=1}^{\eta_{y,u}} \gamma_j \mathbf{a}^*_{\mathrm{UE}}(\phi_{x,u}) \mathbf{a}_{\mathrm{UE}}(\phi_{j,y,u}) \mathbf{a}^*_{\mathrm{BS}}(\theta_{j,y,u}) \mathbf{a}_{\mathrm{BS}}(\theta_{y,w}) ||^2,$$

Thus,

$$I_{y,0} = \frac{\mathrm{G}L(y,0)^{-1}}{\eta_{y,u}U_y} \sum_{w \in \mathcal{U}_y} || \sum_{j=1}^{\eta_{y,u}} \gamma_j \chi_{j,w} ||^2,$$

where $\chi_{j,w}$ is given by,

$$\chi_{j,w} = \begin{cases} 1 & \text{if } \phi_{x,u} = \phi_{j,y,u} \text{ and } \theta_{y,w} = \theta_{j,y,u} \\ \rho_{\mathrm{BS}} & \text{if } \phi_{x,u} = \phi_{j,y,u} \text{ and } \theta_{y,w} \neq \theta_{j,y,u} \\ \rho_{\mathrm{UE}} & \text{if } \phi_{x,u} \neq \phi_{j,y,u} \text{ and } \theta_{y,w} = \theta_{j,y,u} \\ \rho_{\mathrm{BS}}\rho_{\mathrm{UE}} & \text{otherwise.} \end{cases}$$

Now let us look at the Laplace functional of this interference power.

$$L_{I_0}(s) = \mathbb{E}\left[ \exp\left( -s \sum_{y \in \Phi_{\mathrm{BS}}, y \neq x} I_{y,0} \right) \right]$$

$$= \mathbb{E}\left[\exp\left(-s\sum_{y\in\Phi_{\mathrm{BS}},y\neq x}\frac{\mathrm{G}L(y,0)^{-1}}{\eta_{y,0}U_y}\sum_{w\in\mathcal{U}_y}||\sum_{j=1}^{\eta_{y,0}}\gamma_j\chi_{j,w}||^2\right)\right]$$
$$\overset{(a)}{=} \mathbb{E}\left[\prod_{y\in\Phi_{\mathrm{BS}},y\neq x}\mathbb{E}_{\chi_{(.,.)},\gamma_{(.)}}\left[\exp\left(\frac{-s\mathrm{G}L(y,0)^{-1}}{\eta_{y,0}U_y}\sum_{w\in\mathcal{U}_y}||\sum_{j=1}^{\eta_{y,0}}\gamma_j\chi_{j,w}||^2\right)\right]\right].$$

where (a) follows since $\chi$ and $\gamma$ have distributions independent of location $y$. Finding the exact distribution from this expression is intractable. The main bottleneck is that the small scale fading random variables $\gamma_j$, are together clubbed in a single norm expression and thus, although these random variables are assumed to be independent, the distribution of the norm squared for different users in $\mathcal{U}_y$ are correlated exponential random variables. We, thus, find upper and lower bounds in this work.

#### 2.9.4.1 Upper Bound on the Laplace Functional

In order to find an upper bound, the following fact is used. $\chi_{j,w} \geq \rho_{\mathrm{BS}}\rho_{\mathrm{UE}}$. Thus,

$$L_{I_0}(s) \leq \mathbb{E}\left[\prod_{y\in\Phi_{\mathrm{BS}},y\neq x}\mathbb{E}_{\chi_{(.,.)},\gamma_{(.)}}\left[\exp\left(-\frac{s\mathrm{G}\rho_{\mathrm{BS}}^2\rho_{\mathrm{UE}}^2}{L(y,0)\eta_{y,0}U_y}\sum_{w\in\mathcal{U}_y}||\sum_{j=1}^{\eta_{y,0}}\gamma_j||^2\right)\right]\right]$$
$$\overset{(a)}{=} \mathbb{E}\left[\prod_{y\in\Phi_{\mathrm{BS}},y\neq x}\mathbb{E}_{\Xi}\left[\exp\left(\frac{-s\mathrm{G}\rho_{\mathrm{BS}}^2\rho_{\mathrm{UE}}^2\Xi}{L(y,0)\eta_{y,0}}\right)\right]\right]$$
$$= \mathbb{E}\left[\prod_{y\in\Phi_{\mathrm{BS}},y\neq x}\frac{1}{1+s\mathrm{G}L(y,0)^{-1}\rho_{\mathrm{BS}}^2\rho_{\mathrm{UE}}^2}\right],$$

where $\Xi$ is an exponential random variable with mean $\eta_{y,0}$ in (a). In order to find the SINR distribution, we are interested in Laplace functional conditioned

on path loss to serving BS. Thus, conditioning on $L(x,0) = l$ and displacing the points in $\Phi$ to $\mathcal{N}$, similar to Appendix 2.9.3 we get,

$$L_{I_0,l}(s) \leq \mathbb{E}\left[\prod_{t\in\mathcal{N},t\geq l} \frac{1}{1+s\mathrm{G}t^{-1}\rho_{\mathrm{BS}}^2\rho_{\mathrm{UE}}^2}\right]$$
$$= \exp\left(-\int_l^\infty \frac{\Lambda(\mathrm{d}t)}{1+\frac{1}{s\mathrm{G}t^{-1}\rho_{\mathrm{BS}}^2\rho_{\mathrm{UE}}^2}}\right).$$

#### 2.9.4.2 Lower Bound on the Laplace Functional

One obvious lower bound can be obtained using $\chi_{j,w} = 1$. The Laplace functional in this case is the same as for the upper bound with $\rho_{\mathrm{BS}}^2\rho_{\mathrm{UE}}^2$ replaced by 1. However, with the narrow beamwidth for a large number of antennas, this approximation is clearly very pessimistic. We can get a tighter lower bound using the Cauchy-Schwarz inequality as follows.

$$L_{I_0}(s) \geq \mathbb{E}\left[\prod_{y\in\Phi,y\neq x} \mathbb{E}_{\chi_{(.,.)},\gamma_{(.)}}\left[\exp\left(\frac{-s\mathrm{G}L(y,0)^{-1}}{\eta_{y,0}U_y}\left(\sum_{j=1}^{\eta_{y,0}}||\gamma_j||^2\right)\sum_{w\in\mathcal{U}_y}\sum_{j=1}^{\eta_{y,0}}\chi_{j,w}^2\right)\right]\right]$$
$$= \mathbb{E}\left[\prod_{y\in\Phi,y\neq x} \mathbb{E}_{\chi_{(.,.)}}\left[\left(1+\frac{s\mathrm{G}L(y,0)^{-1}}{\eta_{y,0}U_y}\sum_{w\in\mathcal{U}_y}\sum_{j=1}^{\eta_{y,0}}\chi_{j,w}^2\right)^{-\eta_{y,0}}\right]\right].$$

Simplifying the term

$$\Psi_y = \mathbb{E}_{\eta_{y,0}}\left[\left(1+\frac{s\mathrm{G}L(y,0)^{-1}}{\eta_{y,0}U_y}\sum_{w\in\mathcal{U}_y}\sum_{j=1}^{\eta_{y,0}}\chi_{j,w}^2\right)^{-\eta_{y,0}}\right],$$

we get $\Psi_y =$

$$\sum_{i=1}^{\mathrm{N_{BS}}}\sum_{m=0}^{\eta_{y,0}}\binom{\eta_{y,0}}{m}q_{\mathrm{UE},i}^{m+1}(1-q_{\mathrm{UE},i})^{\eta_{y,0}-m}\times$$

$$\sum_{k_1,\ldots,k_{U_y}=1}^{\mathrm{N_{BS}}} \sum_{j_1,\ldots,j_{U_y}=0}^{\eta_{y,0}} \prod_{n=1}^{U_y} \binom{\eta_{y,0}}{j_n} q_{\mathrm{BS},k_n}^{j_n+1} (1-q_{\mathrm{BS},k_n})^{\eta_{y,0}-j_n} \times$$
$$\left(1 + \frac{s\mathrm{G}L(y,0)^{-1}}{\eta_{y,0}U_y} \sum_{j=1}^{\eta_{y,0}} \left( \sum_{n=1}^{U_y} \left(\rho_{\mathrm{BS}}^2 + \mathbb{1}(j \le j_n)\right)(1-\rho_{\mathrm{BS}}^2) \right) \times \right.$$
$$\left. \left(\rho_{\mathrm{UE}}^2 + \mathbb{1}(j \le m)(1-\rho_{\mathrm{UE}}^2)\right) \right)^{-\eta_{y,0}}.$$

The above expression boils down to Lemma 6, for a single path channel. This expression can be further simplified assuming equally probable virtual angles,

$$\Psi_y = \sum_{m=0}^{\eta_{y,0}} \binom{\eta_{y,0}}{m} \left(\frac{1}{\mathrm{N_{UE}}}\right)^m \left(1-\frac{1}{\mathrm{N_{UE}}}\right)^{\eta_{y,0}-m} \sum_{j_1,\ldots,j_{U_y}=0}^{\eta_{y,0}} \times$$
$$\prod_{n=1}^{U_y} \binom{\eta_{y,0}}{j_n} \left(\frac{1}{\mathrm{N_{BS}}}\right)^{j_n} \left(1-\frac{1}{\mathrm{N_{BS}}}\right)^{\eta_{y,0}-j_n} \left(1 + \frac{s\mathrm{G}L(y,0)^{-1}}{\eta_{y,0}U_y} \times \right.$$
$$\left. \sum_{j=1}^{\eta_{y,0}} \left( \sum_{n=1}^{U_y} \left(\rho_{\mathrm{BS}}^2 + \mathbb{1}(j \le j_n)\right)(1-\rho_{\mathrm{BS}}^2) \right) \left(\rho_{\mathrm{UE}}^2 + \mathbb{1}(j \le m)(1-\rho_{\mathrm{UE}}^2)\right) \right)^{-\eta_{y,0}}.$$

Now separating the LOS and NLOS terms and using the Displacement theorem as for the upper bound, the Laplace functional can be given as

$$L_{I_0,l}(s) \ge \exp\left(-\int_l^\infty (1-\mathbb{E}\left[\Psi_{t,\mathrm{L}}\right])\Lambda_{\mathrm{L}}(\mathrm{d}t)\right) \exp\left(-\int_l^\infty (1-\mathbb{E}\left[\Psi_{t,\mathrm{N}}\right])\Lambda_{\mathrm{N}}(\mathrm{d}t)\right).$$

where $\Psi_{t,\mathrm{j}}$ is same as $\Psi_y$ with $y$ replaced by $t$ and $\eta_{y,0}$ replaced by $\eta_{\mathrm{j}}$, for $j \in \{\mathrm{L},\mathrm{N}\}$.

# Chapter 3

# Correction Factor for Analysis of MIMO Wireless Networks With Highly Directional Beamforming[5]

This chapter reconsiders a popular received signal power model, which has been used in most system capacity evaluation studies in mmWave networks until now. Although we do not study new system design issues for mmWave in the chapter, we wish to bring to the notice of the academic community a required modification to an important modeling assumption and it's consequences on system design. In system level analysis for computing coverage and rate performance of wireless networks on $\mathbb{R}^2$ a popular model to compute the received signal power at $X \in \mathbb{R}^2$ from a transmitter (serving/interfering) at $Y \in \mathbb{R}^2$ is as follows [34, 79–81].

$$P_r = P_t \ell(||X - Y||) h G_t(\theta) G_r(\phi), \tag{3.1}$$

[5]This chapter reproduces the content of the following publication. M. N. Kulkarni, E. Visotsky and J. G. Andrews, "Correction factor for analysis of MIMO wireless networks with highly directional beamforming", in *IEEE Wireless Communication Letters*, to appear, 2018. The research performed in this chapter including formulation of the analytical problem, and solving it are primarily my contribution. My co-author, E. Visotsky, helped me in identifying the problem and in generating Fig. 3.1. My co-author, J. G. Andrews, gave regular feedback while I was working on the theoretical results, and while I was writing the paper.

where $P_t$ is the transmit power, $\ell(.)$ is the path loss, $h$ is the small scale fading, $G_t(\theta)$ is the transmit antenna gain and $G_r(\phi)$ is the receive antenna gain. If at all blockage effects are explicitly incorporated in the analysis by differentiating line of sight (LOS) and non-LOS (NLOS) links, then only $h$ and $\ell$ are modeled differently for LOS and NLOS [34]. The antenna patterns $G_t(.)$ and $G_r(.)$ are considered to have the same distribution for LOS/NLOS links. In this work, we will show the importance of incorporating an additional blockage dependent factor in the received signal power when the antenna patterns have very narrow beamwidths and large gains – for example, an antenna pattern having 36 dB gain and $12^o$ half power beamwidth in azimuth. Our analytical model shows that if there are large number of antennas at the transmitter and receiver, which employ analog beamforming, then the additional factor (called as the correction factor) is much less than 1 for NLOS service links but is close to 1 for LOS service links, and equal to 1 for NLOS/LOS interfering links. Such a factor cannot be incorporated by modifying either $h$ or $\ell(.)$ for analyzing signal to interference plus noise ratio (SINR) in highly directional MIMO wireless networks, especially cellular networks, and an example to explain this is given in the Appendix.

Most prior analyses of MIMO wireless networks computing coverage and rate performance with highly directional single user beamforming incorporates a received signal power similar to (3.1) and do not model a channel with LOS/NLOS dependent rank [3, 20, 34, 80–84][6], which gives rise to

[6]Except our prior work in [1], which is detailed in Chapter 2, to the best of our knowledge.

the needed correction factor as we will show in this work. The purpose of this chapter is to make the growing research community using received power model similar to (3.1) aware of the significance of how different rank of the MIMO channel for LOS and NLOS can affect the effective antenna gain and thus the design insights. We will formally define effective antenna gain in this work. Also we propose a quick way to preserve the existing analyses by multiplication of a LOS/NLOS dependent constant for service links but not the interfering links. The example in Appendix is indicative of how this can be done. The correction factor is especially important for analysis of millimeter wave (mmWave) cellular networks, wherein inclusion of blockage effects is crucial and the beamforming is highly directional [34]. All prior works which studied different system design issues in these networks like [3, 20, 82–84] use the received power model in (3.1) without incorporating the correction factor. In Section V, we discuss key implications on system design resulting from incorporation of such a factor.

The analysis in this work is for analog beamforming implementation done at the transmitter and the receiver under consideration. Our analysis along with the simulation results considering a more detailed wideband 3GPP channel model suffice to motivate the inaccuracy of the popular model in (3.1) when the transmit and receive beams are narrow and with large gains. However, more detailed analysis is needed in the future to estimate the correction factor more accurately.

## 3.1 System Model

We concentrate on a single transmitter-receiver pair in a wireless network. $N_t$ and $N_r$ denote the number of transmit and receive antennas. If the link is NLOS, the narrowband channel between the transmitter-receiver is given by [5, 24]

$$\mathbf{H}_{\text{NLOS}} = \kappa \sqrt{\frac{\ell(d)}{\eta}} \sum_{i=1}^{\eta} \gamma_i \mathbf{a}_r(\phi_i) \mathbf{a}_t^*(\theta_i), \tag{3.2}$$

where $\ell(d)$ is the path gain (assumed deterministic function of $d$ for simplicity), $\eta$ is the number of paths (assumed constant), $d$ is the transmission distance in meters and $\gamma_i$ is the small scale fading on path $i$ (random variable such that $\mathbb{E}\left[|\gamma_i|^2\right] = 1$) and $\kappa$ is a normalizing constant such that $\mathbb{E}\left[||\mathbf{H}_{\text{NLOS}}||_F^2\right] = N_t N_r \ell(d)$.

Assuming a uniform linear array at the receiver, the array response vector $\mathbf{a}_r$ is given as $\mathbf{a}_r(\phi_i) = \left[1 \; e^{-j\phi_i} \; e^{-2j\phi_i} \; \ldots \; e^{-(N_r-1)\phi_i}\right]^T$, where $j$ is square root of $-1$. Similarly, one can define $\mathbf{a}_t$ by replacing $N_r$ with $N_t$. Note that $\phi_i$ and $\theta_i$ are spatial angles of arrival and departure (AOA/AOD). It is assumed that these AOAs and AODs are continuous random variables and no assumption on their distribution is made.

If the link is LOS, the narrowband channel is given by [5]

$$\mathbf{H}_{\text{LOS}} = \sqrt{\ell(d)} \left( \sqrt{\frac{\text{K}_\text{R}}{\text{K}_\text{R}+1}} \mathbf{a}_r(\phi_0) \mathbf{a}_t^*(\theta_0) + \kappa \sqrt{\frac{1}{\eta(\text{K}_\text{R}+1)}} \sum_{i=1}^{\eta} \gamma_i \mathbf{a}_r(\phi_i) \mathbf{a}_t^*(\theta_i) \right), \tag{3.3}$$

where $\text{K}_\text{R}$ is the Rician K-factor. AOA and AOD given by $\phi_0$ and $\theta_0$ are constants corresponding to the direct LOS path between the receiver and

the transmitter. Rest of the AOA/AOD are continuous random variables. $\eta$ and $\kappa$ could have different LOS-specific values here, as compared to (3.2).

Assuming $\mathbf{w}$ is the combiner employed by the receiver and $\mathbf{f}$ is the precoder employed by the transmitter, the received signal power model is given as $P_r^{\text{multi}} = ||\mathbf{w}^*\mathbf{H}\mathbf{f}||^2$, where $\mathbf{H}$ is the $N_r \times N_t$ channel which could be either $\mathbf{H}_{\text{LOS}}$ or $\mathbf{H}_{\text{NLOS}}$. We constrain $\mathbf{w}$ and $\mathbf{f}$ to be chosen of the form $\frac{1}{\sqrt{N_r}}\mathbf{a}_r(.)$ and $\frac{1}{\sqrt{N_t}}\mathbf{a}_t(.)$, respectively, which is basically employing analog beamforming using phase shifters at both the receiver and the transmitter. If the transmitter-receiver pair form a desired communication link, $\mathbf{w}$ and $\mathbf{f}$ are chosen so as to maximize $P_r^{\text{multi}}$. If the transmitter-receiver pair form an interfering link, then $\mathbf{w}$ and $\mathbf{f}$ can be arbitrary.

Most analytical studies to compute coverage and rate performance cannot afford to use the received signal power model defined above for tractability. As mentioned in Section I, a simplified model similar to (3.1) is generally used. Now we will define a generative model for such a simplified model. We define a keyhole channel as follows [85]. $\mathbf{H}_{\text{keyhole}} = \sqrt{\ell(d)}\gamma\mathbf{a}_r(\phi)\mathbf{a}_t^*(\theta)$, where $\mathbb{E}\left[|\gamma|^2\right] = 1$ and $\{\theta, \phi\}$ could have arbitrary distribution. Now, $P_r^{\text{keyhole}}$ is defined as $||\mathbf{w}^*\mathbf{H}_{\text{keyhole}}\mathbf{f}||^2$. If the transmitter-receiver pair is a desired signal link, $\mathbf{w} = \frac{1}{\sqrt{N_r}}\mathbf{a}_r(\phi)$ and $\mathbf{f} = \frac{1}{\sqrt{N_t}}\mathbf{a}_t(\theta)$ to maximize $P_r^{\text{keyhole}}$ and thus, $P_r^{\text{keyhole}} = |\gamma|^2\ell(d)N_tN_r$. If the transmitter-receiver pair is an interfering link with $\mathbf{w} = \frac{1}{\sqrt{N_r}}\mathbf{a}_r(\phi')$ and $\mathbf{f} = \frac{1}{\sqrt{N_t}}\mathbf{a}_t(\theta')$ for some arbitrary $\phi'$ and $\theta'$, then $P_r^{\text{keyhole}} = \ell(d)|\gamma|^2G_r(\phi, \phi')G_t(\theta', \theta)$, where $G_r(\phi, \phi') = ||\frac{1}{\sqrt{N_r}}\mathbf{a}_r^*(\phi)\mathbf{a}_r(\phi')||^2$. Similarly $G_t$ can be written replacing subscript $r$ with $t$ and $\phi$ with $\theta$. Un-

like the desired signal power case, here $\phi'$ and $\theta'$ are not chosen to maximize $||\mathbf{w}^*\mathbf{H}_{\text{keyhole}}\mathbf{f}||^2$ but can be random angles distributed according to some continuous distribution.

We wish to compare $\mathbb{E}\left[P_r^{\text{multi}}\right]$ with $\mathbb{E}\left[P_r^{\text{keyhole}}\right]$. This comparison will highlight how important it is to consider rank $> 1$ channels for LOS and NLOS in terms mean received signal power since the keyhole channel is always rank 1. In order to quantify this comparison, we define a correction factor as follows.

**Definition 3.** The proposed *correction factor* to estimate the received signal power on a serving/interfering link is defined as $\Upsilon = \mathbb{E}\left[P_r^{\text{multi}}\right] / \mathbb{E}\left[P_r^{\text{keyhole}}\right]$.

Note that for serving links $\mathbb{E}\left[P_r^{\text{keyhole}}\right] = N_t N_r \ell(d)$ irrespective of LOS/NLOS as per our analytical model.

**Definition 4.** The *effective antenna gain* is defined as the actual received signal power (on serving/interfering links) normalized by the path loss and the transmit power of the signal.

Note that the effective antenna gain is in general a random variable. As per our analytical model, it is equal to $P_r^{\text{multi}}/\ell(d)$. Considering our system model, wherein $\ell(d)$ is deterministic the mean effective antenna gain for a serving link is given by $\Upsilon N_t N_r$, where $\Upsilon$ is the correction factor for a serving link. Our proposal is that if one wants to use a simplified received power model like in (3.1) for system level analysis, wherein the impact of beamforming is captured through a spatial gain pattern at the transmitter and receiver, then

the corrected received signal power on serving and interfering links is obtained by multiplying $\Upsilon$ to the estimate in (3.1). Since here a keyhole model is used to generate the simplified received power model in (3.1), the corrected received signal power is $\Upsilon P_r^{\text{keyhole}}$.

## 3.2 Computing $\Upsilon$ When $N_t, N_r$ Are Large

Before we state the results, we make a quick observation based on the result in [68].

**Observation 1:** As $N_r \to \infty$ and $N_t \to \infty$, the left singular vectors corresponding to non-zero singular values of (3.2) and (3.3) converge in chordal distance to $\frac{1}{\sqrt{N_r}}\mathbf{a}_r(\phi_i)$, with $i = 1\ldots,\eta$ for (3.2) and $i = 0,\ldots,\eta$ for (3.3). Similarly, the right singular vectors corresponding to non-zero singular values of (3.2) and (3.3) converge to $\frac{1}{\sqrt{N_t}}\mathbf{a}_t(\phi_i)$.

**Observation 2:** As $N_r \to \infty$, $\mathbf{a}_r^*(\phi_i)\mathbf{a}_r(\phi_j)/N_r \to \mathbb{1}(i=j)$. Similarly $\mathbf{a}_t^*(\theta_i)\mathbf{a}_t(\theta_j)/N_t \to \mathbb{1}(i=j)$ as $N_t \to \infty$.

**Theorem 4.** *Large $N_t$ and $N_r$ is assumed. If the link is a NLOS service link, then* $\mathbb{E}\left[P_r^{\text{multi}}\right] \approx N_t N_r \ell(d) \times \mathbb{E}\left[\max_{i=1,\ldots,\eta} |\gamma_i|^2\right]/\eta$.

*Proof.* Optimal combiner and precoder correspond to the singular vectors corresponding to the maximum singular value norm of the channel matrix. Making use of Observation 1 for NLOS channel with large number of antennas, $\mathbf{w} = \frac{1}{\sqrt{N_r}}\mathbf{a}_r(\phi_1)$ and $\mathbf{f} = \frac{1}{\sqrt{N_t}}\mathbf{a}_t(\theta_1)$ assuming $|\gamma_1| = \max_i |\gamma_i|$, without loss of

generality. Thus,

$$P_r^{multi} = ||\mathbf{w}^* \mathbf{H}_{\text{NLOS}} \mathbf{f}||^2 = \left|\left| \kappa \sqrt{\frac{\ell(d)}{\eta N_t N_r}} \sum_{i=1}^{\eta} \gamma_i \mathbf{a}_r^*(\phi_1) \mathbf{a}_r(\phi_i) \mathbf{a}_t^*(\theta_i) \mathbf{a}_t(\phi_1) \right|\right|^2 .$$

Note that Observation 1 implies that the non-zero singular values of (3.2) are given by $\sqrt{\frac{\ell(d) N_t N_r}{\eta}} \kappa \gamma_i$. Thus, $||\mathbf{H}_{\text{NLOS}}||_F^2 = \kappa^2 \ell(d) N_t N_r \sum_{i=1}^{\eta} |\gamma_i|^2 / \eta$, which is computed using the fact that square of Frobenius norm equals sum of squares of singular values of a matrix. Thus $\mathbb{E}\left[||\mathbf{H}_{\text{NLOS}}||_F^2\right] = N_t N_r \ell(d) \kappa^2$, which implies that the normalizing constant $\kappa = 1$. Similarly, $\kappa = 1$ in (3.3).

Since the AODs/AOAs are continuous random variables, any two such angles are unequal with probability 1. Using the orthogonality of the array response vectors for unequal AODs/AOAs, we get $P_r^{\text{multi}} \approx \left|\left| N_r N_t \gamma_1 \sqrt{\frac{\ell(d)}{\eta N_t N_r}} + 0 \right|\right|^2 = N_t N_r \ell(d) \frac{|\gamma_1|^2}{\eta}$ with probability 1. Thus, the expectation of $P_r^{\text{multi}}$ is $N_t N_r \ell(d) \frac{\mathbb{E}\left[|\gamma_1|^2\right]}{\eta}$. The result is approximate as we used asymptotic results in Observations 1 and 2 for finite number of antennas.

For LOS, since $\mathbb{E}\left[|\gamma_i|^2\right] = 1$, by Markov inequality $\mathbb{P}\left(|\gamma_i|^2 > \eta \mathrm{K_R}\right) < 1/\eta \mathrm{K_R}$. Thus, owing to $\mathrm{K_R} \gg 1$ with high probability the maximum singular value corresponds to the direct LOS path. This implies that $\mathbf{w} = \frac{1}{\sqrt{N_r}} \mathbf{a}_r(\phi_0)$ and $\mathbf{f} = \frac{1}{\sqrt{N_t}} \mathbf{a}_t(\theta_0)$, which are singular vectors corresponding to the maximum singular value as per Observation 1. Thus, it is concluded that

$$P_r^{\text{multi}} \approx ||\mathbf{a}_r^*(\phi_0) \mathbf{H}_{\text{LOS}} \mathbf{a}_t(\theta_0)||^2 = \left|\left| \sqrt{\ell(d) N_t N_r \mathrm{K_R} / (1 + \mathrm{K_R})} + \rho \right|\right|^2 ,$$

where

$$\rho = \frac{1}{\sqrt{\eta(\mathrm{K_R} + 1)}} \sum_{i=1}^{\eta} \gamma_i \mathbf{a}_r^*(\phi_0) \mathbf{a}_r(\phi_i) \mathbf{a}_t^*(\theta_i) \mathbf{a}_t(\theta_0).$$

Note that using Observation 2, we have $\rho \approx 0$ by similar arguments as for NLOS case considering the angles of arrival/departure are continuous random variables. Thus, $\mathbb{E}\left[P_r^{\text{multi}}\right] \approx N_t N_r \ell(d) \frac{\mathrm{K_R}}{\mathrm{K_R}+1}$. □

**Corollary 3.** *Large $N_t$ and $N_r$ is assumed and the link under consideration is assumed to be a service link. If $\gamma_i$ are complex normal random variables and independent of each other, $\Upsilon \approx \frac{1}{\eta}\sum_{k=1}^{\eta}(1/k)$ if the link is NLOS. If $\gamma_i$ are all identical to complex normal $\gamma_1$, $\Upsilon \approx \frac{1}{\eta}$ for NLOS link. For LOS link and $\mathrm{K_R} \gg 1$, $\Upsilon \approx \frac{\mathrm{K_R}}{1+\mathrm{K_R}} \approx 1$.*

*Proof.* If $\gamma_i$ are complex normal random variables, $|\gamma_i|^2$ are exponentially distributed with unit mean. Also these are independent random variables. Thus, $\mathbb{E}\left[\max_{i=1,\ldots,\eta}|\gamma_i|^2\right] = \sum_{k=1}^{\eta}(1/k)$ [86]. By Theorem 4 and $\mathbb{E}\left[P_r^{\text{keyhole}}\right] = N_t N_r \ell(d)$, $\Upsilon \approx \mathbb{E}\left[\max_i |\gamma_i|^2\right]/\eta = \frac{1}{\eta}\sum_{k=1}^{\eta}(1/k)$ if $\gamma_i$ are complex normal random variables and independent of each other. Similarly, the other two results are derived. □

From Corollary 3, NLOS received signal power can be significantly overestimated with the keyhole model for $\eta = 10$, which translates to $\Upsilon = -4.6$dB if $\gamma_i$ are identically equal to $\gamma$, and to $\Upsilon = -10$dB if $\gamma_i$ are independently but identically distributed. Note that this is an analytical result and that well accepted wideband models (like in [5]) will have unequal distribution of powers amongst paths within and across different clusters. Estimating $\Upsilon$ in these settings is an avenue for further research.

**Theorem 5.** *Let the transmitter and receiver beamforming vectors be* $\frac{\mathbf{a}_t(\theta')}{\sqrt{N_t}}$ *and* $\frac{\mathbf{a}_r(\phi')}{\sqrt{N_r}}$. *If* $\gamma_i$ *are independent zero mean random variables with unit variance,* $\{\theta_i\}$ *are identically distributed to* $\theta$, $\{\phi_i\}$ *are identically distributed to* $\phi$ *and* $\{\gamma_i\}$ *are independent of all AOAs/AODs, then* $\mathbb{E}\left[P_r^{\text{multi}}\right] = \mathbb{E}\left[P_r^{\text{keyhole}}\right]$ *for NLOS interfering links.*

*Proof.* The received signal power considering a multipath channel in (3.2) is given by

$$P_r^{\text{multi}} = \frac{\ell(d)}{\eta}\left|\left|\sum_{i=1}^{\eta}\frac{\gamma_i}{\sqrt{N_t N_r}}\mathbf{a}_r^*(\phi')\mathbf{a}_r(\phi_i)\mathbf{a}_t^*(\theta_i)\mathbf{a}_t(\theta')\right|\right|^2. \tag{3.4}$$

Using independence of $\gamma_i$ and that these are zero mean random variables, the cross terms while expanding the norm squared in (3.4) become zero and thus, $\mathbb{E}\left[P_r^{\text{multi}}\right]$ is equal to

$$\frac{\ell(d)}{\eta}\sum_{i=1}^{\eta}\mathbb{E}\left[|\gamma_i|^2\right]\mathbb{E}\left[G_r(\phi',\phi_i)G_t(\theta_i,\theta')\right] = \frac{\ell(d)\eta\mathbb{E}\left[G_r(\phi',\phi_1)G_t(\theta_1,\theta')\right]}{\eta} = \mathbb{E}\left[P_r^{\text{keyhole}}\right].$$

□

Theorem 5 indicates that a correction factor is not necessary for NLOS interfering links, if the assumptions in the theorem hold true. A result of similar nature can be stated for LOS interfering links. However, depending on the structure of the arrays, the per-element antenna gains and joint distribution of $\{\gamma_i, \phi_i, \theta_i\}$ a non-unity correction factor may be necessary for interfering links. Next, we will validate the need for a correction factor with some simulations using the 3GPP NR channel model [5].

## 3.3 Simulation Result with 3GPP model

We consider two MIMO systems with link lengths 100 meters operating at 73 GHz carrier frequency. One is LOS and the other is NLOS. $8 \times 8$ uniform planar array with half wavelength spacing is assumed at the transmitters and the receivers. Note that considering a $8 \times 8$ antenna array system is realistic for mmWave backhaul networks wherein both ends of a communication link are base stations (BSs) [16]. Effective antenna gain for each of these MIMO systems is computed as $P_r^{\text{multi}}/\ell(d)$ as per Definition 3. Here, $P_r^{\text{multi}}$ was computed considering the 3GPP NR channel model [5] along with optimal precoders and combiners that maximize the SNR and a unit transmit power. Several realizations of the 3GPP channel were simulated for both the links. The distribution of effective antenna gain seen by the LOS and NLOS link is plotted in Fig. 3.1. As seen from Fig. 3.1 there is a drop of about 12 dB in NLOS median gain compared to LOS, which is very significant. The implication of such drop in effective antenna gain is discussed in next section. The LOS effective antenna gain in Fig. 3.1 is very close to $10\log_{10}(64 \times 64) = 36\text{dB}$ , as expected, since correction factor for LOS links is negligible as per our analysis. Surprisingly the drop in NLOS gain is equal to $-10\log_{10} 19$, wherein 19 is the mean number of NLOS clusters in the 3GPP model. This equals our analytical estimate of $1/\eta$ considering $\eta = 19$. A more accurate analysis explicitly modeling different clusters with multiple rays and correlated small scale fading is a possible future work.

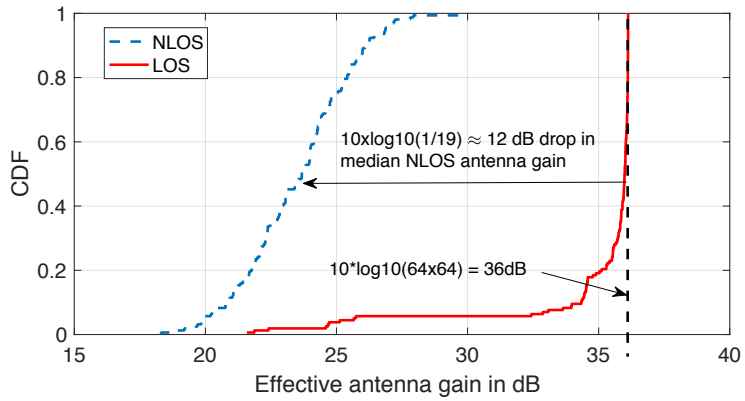



Figure 3.1: Comparison of effective antenna gain for LOS and NLOS links with new radio 3GPP channel model [5].

## 3.4 Implications and Applicability of the Work

This work is applicable for MIMO wireless networks with highly directional single stream beamforming at the transmitter and the receiver. The analysis can also be extended for multi-user MIMO with large number of transmit and receive antennas. In short, whenever the underlying signal processing of a large MIMO system is abstracted to compute the received signal power as a product of a single input single output (SISO) received signal power and some spatial antenna gain patterns at the transmitter/receiver for simplified analysis, there will be a need for incorporating the correction factor to make sure that identical antenna gain patterns are not multiplied for LOS and NLOS links, as well as serving and interfering links. The implications of the work are prominent in the following scenarios. For dense outdoor-to-outdoor cellular networks, a user would likely associate with a LOS BS and thus the signal to noise ratio (SNR) coverage estimates wouldn't vary significantly, except for

the tail probability when a user associates with a NLOS BS that affects the cell edge rates. Otherwise, we expect such a correction factor to be significant since there is significant probability of connecting to a NLOS BS since the SNR distribution itself will shift by $\Upsilon$. We expect the significance of such a correction factor to also be significant in analysis of multi-hop mmWave cellular networks wherein the fiber site deployment will be relatively sparse and thus there will be a question as to whether a relay should go for a NLOS direct hop to fiber base station or whether it should relay over multiple LOS hops. Given that the correction factor introduced in this letter doesn't affect LOS links but strongly affects NLOS links, LOS hops will be even more strongly favoured over NLOS hops. Neglecting the correction factor but using a model like (3.1) can lead to misleading insights.

## 3.5 Summary

In this chapter, we suggested modifications in a popular simplified received signal power model with single stream beamforming employed by the transmitter and the receiver in the regime when the beams have high gain and narrow beamwidth. Based on our analytical results as well as supporting simulations we confirmed the importance of incorporating the suggested modifications in system level analysis of MIMO wireless networks. In short, whenever the underlying signal processing of a large MIMO system is abstracted to compute the received signal power as a product of a single input single output (SISO) received signal power and some spatial antenna gain pat-

terns at the transmitter/receiver for simplified analysis, there will be a need for incorporating the proposed correction factor to make sure that identical antenna gain patterns are not multiplied for LOS and NLOS links, as well as serving and interfering links.

## 3.6 Appendix: An Example Demonstrating The Use Of Correction Factor

Consider a receiving user at origin and a collection of transmitting base stations (BSs) in $\mathbb{R}^2$ including a BS at $Y \in \mathbb{R}^2$ which is NLOS with respect to the user. We want to understand the SINR performance of that user using the simplified received signal power in (3.1) that models beamforming through a spatial antenna pattern. Our proposal is to introduce the correction factor to compute the received signal powers. In principle, this factor is different for LOS and NLOS as well as for service and interfering links. For simplicity of exposition, we will consider the correction factor to be much less than 1 for NLOS serving links and equal to 1 for rest of the cases, which is an outcome of our asymptotic analysis. First to evaluate whether $Y$ is an interferer or a serving BS – usually the serving BS is the one with maximum received signal power averaged over $h$ – one has to multiply a correction factor that is much less than 1 to the received signal power from $Y$ to origin. However, if it is determined that the BS does not serve the user but is a potentially interfering BS, then the correction factor is equal to 1 while computing the interference power from the same BS at $Y$ to the receiver at origin. Such a modification

in received power, which is done differently for service and interfering links cannot be done by modifying $\ell(.)$ or $h$.

# Chapter 4

# Resource Allocation in Self-backhauled Networks with a Single Backhaul Hop[7]

In Chapter 2, the focus was on understanding how to use the large antenna arrays in mmWave cellular networks. In this and the next chapter, the focus is on another key system design issue for the success of mmWave cellular. Since even initial deployments of mmWave cellular need to be dense to provide sufficient coverage, it is necessary to develop cost effective deployment solutions. Self-backhauling offers a simple cost-saving strategy to enable dense millimeter wave cellular networks [21, 46, 87]. A self-backhauled network has two types of base stations (BSs) – master BSs (MBSs) and slave BSs (SBSs). SBSs wirelessly backhaul users' data to/from the fiber backhauled MBSs through either a direct wireless connection or over multiple SBS-SBS hops, sharing the spectrum with access links [39]. A fundamental problem

[7]This chapter reproduces the content of the following publication. M. N. Kulkarni, J. G. Andrews and A. Ghosh, "Performance of dynamic and static TDD in self-backhauled millimeter wave cellular networks", in *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6460-6478, Oct. 2017. The research performed in this chapter including the formulation of the analytical problem, solving it and generating all numerical results are primarily my contribution. My co-authors, J. G. Andrews and A. Ghosh, guided me in identifying the research problem and setting up the system model through several brainstorming sessions. They also gave me regular feedback while I was solving the problem, and while I was writing the paper.

for designing a self-backhauled network is to split the available time-frequency resources between uplink (UL) and downlink (DL) and for the access and backhaul links. In this chapter, a generic random spatial model is developed for studying the resource allocation problem in two hop self-backhauled mmWave cellular networks, with a focus on comparing static and dynamic time division duplexing (TDD) with synchronized or unsynchronized access-backhaul (SAB or UAB).

## 4.1 Dynamic TDD with unsynchronized access-backhaul:– motivation and prior work

Conventionally, a network-wide static split of resources is done between UL and DL, meaning that every BS follows a common UL-DL split of time-frequency resources. Such a static split can be very inefficient in dense networks wherein the load per base station is highly variable, as shown in Fig. 4.1(a). Although the network has overall 50% UL users, the fraction of UL users per BS varies from 16% to 100%, and thus a network wide $50-50$ split between UL and DL resources is wasteful. Dynamic TDD is a class of scheduling schemes wherein every BS is free to choose its own UL-DL split [88, 89]. Widespread use of this TDD scheme was challenging for sub-6GHz networks owing to cross-interference between UL transmissions in one cell and DL transmissions in neighboring cells [88, 89]. Since DL transmissions generally have more power than UL, dynamic TDD generally hurts UL signal to interference plus noise ratio (SINR). At mmWave frequencies, however, dynamic TDD is

expected to perform much better given the likely noise-limited behaviour due to directionality and large bandwidth [21, 40, 90]. Furthermore, the significance of enabling dynamic TDD in future cellular networks is predicted to be even higher for meeting ultra low latency and high throughput requirements of the future wireless technologies [91, 92]. Stochastic geometry has been used to quantify the cross UL-DL interference effects through calculating the *SINR distribution* in sub-6GHz cellular [93], device-to-device enhanced networks [94] and UL mmWave cellular networks [90] but there is no comprehensive UL-DL *rate* analysis with dynamic TDD. In this work, we characterize the gains with dynamic TDD in mmWave cellular networks for UL and DL through explicit mean rate formulas as a function of network parameters and a simple interference mitigation scheme.

Incorporating relays in cellular networks was an afterthought, primarily for coverage enhancement, in current deployments of cellular networks. Two-hop relaying was introduced in 3GPP release 10 [95, Ch. 18]. However, mmWave cellular networks are expected to have dense deployments right from the start to provide sufficient coverage overcoming the enhanced blockage effects and to meet the desired 5G data rates for enabling extreme mobile broadband applications [34, 56]. Thus, a simple cost saving strategy to enable flexible deployments is to have a fraction of BSs wirelessly backhauling data to the rest which have fiber backhaul connectivity, that motivates self-backhauled mmWave cellular networks. Traditionally, in-band implementation of relay networks is restricted to synchronized access-backhaul (SAB), wherein

the access and backhaul links are active on non-overlapping time slots [95, Ch. 18]. However, from resource allocation perspective, an MBS needs more backhaul slots than SBSs in self-backhauled networks. This is not possible with the conventional SAB implementation. An example is shown in Fig. 4.1(b) wherein there are 2 SBSs connected to an MBS. With SAB, the second SBS is silent in a backhaul slot when first SBS is scheduled by the MBS. In fact, the second SBS could have utilized the unscheduled backhaul slots for communicating with its UEs. This issue will be magnified if there are tens or hundreds of SBSs connected to an MBS. An SBS *poaching* the unscheduled backhaul slots for access is said to employ an unsynchronized access-backhaul (UAB) strategy, wherein access and backhaul links need not be scheduled on orthogonal resource blocks. Introducing the above mentioned implementation of UAB, however, comes at a cost of increasing interference on the backhaul links which makes it non-trivial to choose UAB over SAB. Again, the subdued interference effects at mmWave make UAB attractive for practical implementations. UAB has been implicitly incorporated in algorithmic solutions to the resource allocation problem in sub-6GHz relay networks [96] and more recently in mmWave self-backhauled networks [40, 41][8] In this work, we capture the tradeoff between increasing interference and better resource allocation with UAB through our random spatial model, and the analysis can be used to compute optimal *poaching probabilities* (defined in Section 4.3.2) to strike

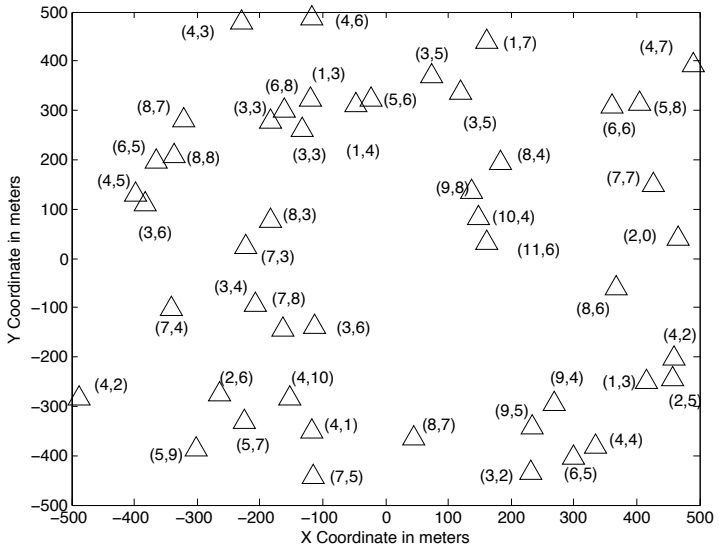
[8]The term "integrated access-backhaul" coined in [41, 97] by Qualcomm and AT&T is same as "UAB" in this work, although our heuristic implementation has a more specific form described in Section 4.3.

a balance. In [46], UAB was implicitly employed, although the focus was on noise-limited mmWave cellular networks. Previous stochastic geometry analysis of relay networks, like in [98–100], did not incorporate UAB and also was focused on sub-6GHz cellular networks.

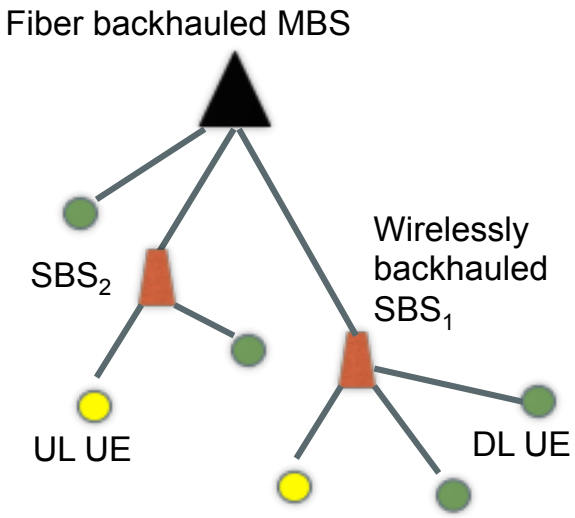
## 4.2 Contributions

**UL and DL analysis of dynamic TDD in mmWave cellular networks.** This is the first work to our knowledge to analyze UL and DL SINR distribution and mean rates in dynamic TDD enabled mmWave cellular networks. A time-slotted system is considered. In a typical access frame, all initial slots are prioritized for DL scheduling and later slots for UL scheduling. Such a prioritization is shown to have inherent UL interference mitigation and the variation of SINR across time slots can be as large as $10 - 15$ dB. This translates to some gain in mean rate as well, but is more crucial for decreasing UL SINR outage probabilities. PPP deployment for users and base stations is assumed for the analysis.

**UL and DL analysis of mmWave self-backhauled cellular networks with unsynchronized access-backhaul.** Achievable mean rates with SAB and UAB are compared in self-backhauled mmWave cellular networks. The optimal number of slots to be exclusively allocated for access is shown to be non-increasing with UAB as compared to SAB. A PPP approximation is proposed and validated for characterizing the interference distribution with UAB, which can have a variety of applications as mentioned in




(a) Dynamic TDD: Varying fraction of UL users per BS. Triangles are BSs and $(x, y)$ are number of UL and DL UEs.




(b) UAB: Need more backhaul slots at MBS than SBS.

Figure 4.1: Motivation for dynamic TDD and UAB.

Section 4.7.

**Engineering insights.** The comparison of resource allocation schemes considered in this chapter are fundamentally dependent on more than ten system parameters, and thus it is not possible to enumerate concrete regimes wherein one strategy will outperform another. Also, dynamic TDD may be the preferred choice over static TDD for DL users but not for UL users, and UAB with no exclusive access slots may be desirable for the users connected to SBSs but not for those connected to MBSs. The analytical formulae provided in this chapter provides a transparent approach to compare the resource allocation schemes for different networks and propagation settings in terms of mean rates and SINR distributions of a typical UL and DL user in the network. Dynamic TDD and UAB *usually* outperform or at least provide similar performance to load aware static TDD and SAB in terms of mean rate of a typical user in millimeter wave cellular networks operating with large bandwidths (order of GHz). The gains of dynamic TDD over static TDD are larger for low load, and asymmetric traffic scenarios. Load aware static TDD can still be preferable over dynamic TDD in interference-limited highly loaded scenarios with symmetric UL and DL traffic requests on an average per BS. We further find that there is no need for asymmetric traffic or low UE load for gains with UAB over SAB and we just need sufficiently large number of SBS per MBS. Self-backhauling is indeed a low cost coverage solution that can enable flexible deployments, but not particularly useful to enhance mean rates if the same antenna array is used by SBSs for both access and backhaul links. Employing

higher spectral efficiency backhaul links is important to harvest the benefits of dynamic TDD and UAB.

## 4.3 System Model

### 4.3.1 Spatial distribution of base stations and users

Let $\Phi_m$ and $\Phi_s$ denote independent PPPs on $\mathbb{R}^2$ of MBSs and SBSs with density $\lambda_m$ and $\lambda_s$ BSs/km$^2$. Let $\Phi_b$ denote the superposition of the two BS PPPs and $\lambda_b = \lambda_m + \lambda_s$ denote its density. User equipments (UEs) are distributed as an independent homogeneous PPP $\Phi_u$ with density $\lambda_u$ UEs/km$^2$ on $\mathbb{R}^2$. A fraction $\eta$ of UEs have DL requests and the rest of them UL. $\Phi_{ul}$ and $\Phi_{dl}$ denote the UL and DL UE point processes with densities $(1-\eta)\lambda_u$ and $\eta\lambda_u$, respectively. UEs always have data to transmit/receive. All devices are half duplex.

### 4.3.2 TDD frames and scheduling

In the following discussion, UL denotes UE to BS links for access and SBS to MBS links for backhaul. Similarly, DL denotes the BS to UE links for access and MBS to SBS for backhaul.

Fig. 4.2(a) shows the TDD frame structure. Each frame consists of 4 subframes for DL access, UL access, DL backhaul, and UL backhaul. There are $\mathrm{F}_{ad}$, $\mathrm{F}_{au}$, $\mathrm{F}_{bd}$, $\mathrm{F}_{bu}$ slots, each of duration T, in the 4 subframes. We denote by $\mathrm{F}_a = \mathrm{F}_{ad} + \mathrm{F}_{au}$, $\mathrm{F}_b = \mathrm{F}_{bd} + \mathrm{F}_{bu}$, and $\mathrm{F} = \mathrm{F}_a + \mathrm{F}_b$. We add a subscript $X$ to each of these to denote the sub-frame size for BS at $X \in \Phi_b$. The terminology

Table 4.1: Notation summary and default numerical parameters

| **Notation** | **Parameter(s)** | **Value(s) if applicable** |
|---|---|---|
| $\Phi_u$,$\Phi_b$, $\Phi_m$,$\Phi_s$ | UE, BS, MBS and SBS PPP on $\mathbb{R}^2$ | – |
| $\lambda_u, \lambda_b$, $\lambda_m, \lambda_s$ | Density of UE, BS, MBS and SBS PPP | 200, 100, 20, 80 (per $\text{km}^2$) |
| $N_u, N_d$, $N_s$ | Number of UL UEs, DL UEs and SBSs. Add subscript $X$ for BS at $X$ | – |
| $X^*$, $X^{**}$ | $X^*$ is BS serving UE at origin and $X^{**}$ is MBS serving $X^* \in \Phi_s$ | – |
| $\mathrm{P}_m, \mathrm{P}_s$, $\mathrm{P}_u$ | Transmit powers | 30, 30, 20 dBm [56] |
| $\Delta_m, \Delta_s$, $\Delta_u$ | Half power beamwidth | $10^o, 10^o, 60^o$ [34, 56] |
| $G_m, G_s$, $G_u$ | Main lobe gain | $24, 24, 6$ dB [24, 34, 87] |
| $g_m, g_s$, $g_u$ | Side lobe gain | $-4, -4, -14$ dB [34] |
| $B_\nu$, $\mathcal{A}_\nu$ | Association bias and probability towards BS of tier $\nu \in \{m, s\}$ | $B_s = B_m = 0$ dB |
| $f_c$, B | Carrier frequency and bandwidth | 28 GHz, 200 MHz |
| $p_{\text{LOS}}$, $D_{\text{LOS}}$ | Blockage parameters | 0.3, 200 m [47] |
| $\alpha_l, \alpha_n$ | LOS, NLOS path loss exponents | 2.1, 3.4 [56] |
| $\mathrm{C}_0$ | 1m reference distance omnidirectional path loss | $\left(3 \times 10^8 / 4\pi f_c\right)^2$ |
| $\sigma^2$ | thermal noise (in dBm) | $-174 + 10\log_{10}(\mathrm{B}) + 5$ |
| $\eta, \delta, \mathrm{F}$ | Fraction of DL UEs, fraction of access slots, frame size | 0.5, 0.5, 1 |
| $\ell, \mu$ | Access/backhaul or LOS/NLOS link | $\ell \in \{a, b\}$, $\mu \in \{l, n\}$ |
| $t$ | Tier of BS PPP | $t \in \{m, s\}$ |
| $i$ | Slot index | $1 \leq i \leq \mathrm{F}$ |
| $w_a, w_b$ | Resource allocation scheme in access and backhaul subframe | $w_a \in \{S, D\}$, $w_b \in \{\text{UAB}, \text{SAB}\}$ |

$i^{\text{th}}$ slot would refer to the $i^{\text{th}}$ slot starting from the beginning of the TDD frame and $i$ varies from 1 to F. We neglect the slots allocated for control signals and subframe switching [101], although this can be incorporated by scaling the rate estimates in this work by a constant factor.

All BSs allocate $\delta$ fraction of F for access and rest for backhaul. If $\delta \mathrm{F} < 1$ then in every time slot a coin is flipped with this probability to decide whether the slot is for access or backhaul, which is synchronously adopted by all BSs. Optimization over $\delta$ is done numerically based on mean rate analysis in Section 4.6. Allowing different BSs to have a different $\delta$ is possible but for analytical tractability we do not consider such a scenario. Thus, $\mathrm{F}_a = \lceil \delta \mathrm{F} \rceil$ with probability $\delta \mathrm{F} - \lfloor \delta \mathrm{F} \rfloor$, and $\mathrm{F}_a = \lfloor \delta \mathrm{F} \rfloor$ otherwise.

Let $\gamma_{\ell,w,X}$ denote the fraction of slots allocated for DL transmissions in subframe of type $\ell \in \{a, b\}$ by BS at location $X$, $w \in \{S, D\}$ denote static and dynamic TDD schemes when $\ell = a$, and $w \in \{\mathrm{SAB}, \mathrm{UAB}\}$ denote synchronized and unsynchronized access-backhaul schemes when $\ell = b$. More on these schemes is discussed in the following text. The above notation implies that $\mathrm{F}_{ad,X} = \lceil \mathrm{F}_a \gamma_{a,w,X} \rceil$ with probability $\mathrm{F}_a \gamma_{a,w,X} - \lfloor \mathrm{F}_a \gamma_{a,w,X} \rfloor$, and $\mathrm{F}_{ad,X} = \lfloor \mathrm{F}_a \gamma_{a,w,X} \rfloor$ otherwise. Similarly for $\mathrm{F}_{bd,X}$ by replacing $\gamma_{a,w,X}$ with $\gamma_{b,w,X}$ and $\mathrm{F}_a$ with $\mathrm{F} - \mathrm{F}_a$.

#### 4.3.2.1 Scheduling in access subframes

We consider the following schemes for choosing $\gamma_{a,w,X}$. In each slot, a BS randomly schedules an UL/DL UE uniformly from the set of connected

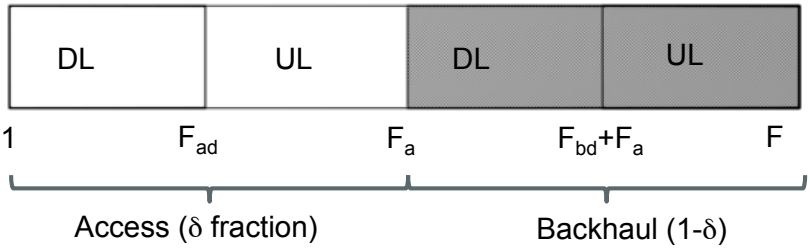
UEs.

- **Static TDD.** Here, $\gamma_{a,S,X} = \gamma_a$, which is a fixed constant independent of $X \in \Phi_b$. This can be a completely load unaware scheme if $\gamma_a$ is irrespective of $\eta$, and a load aware scheme if $\gamma_a$ is dependent on $\eta$. We focus on a load aware scheme wherein $\gamma_a = \eta$.

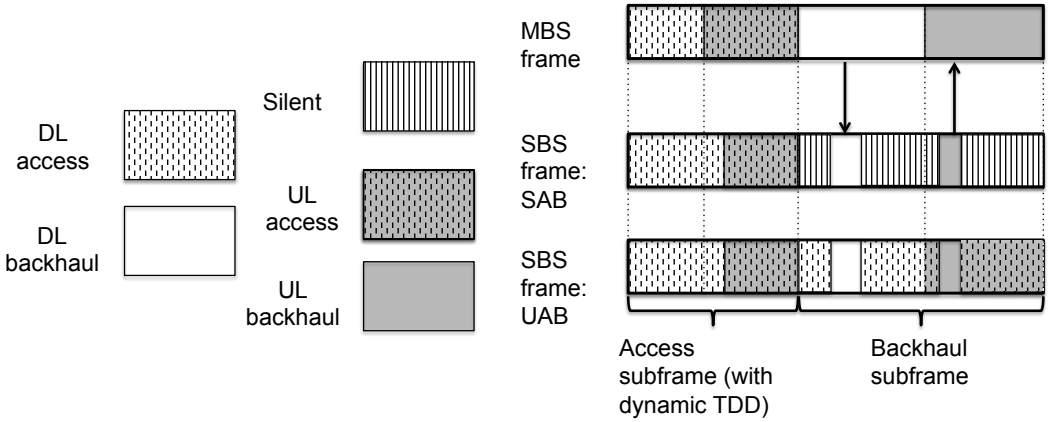
- **Dynamic TDD.** Now, we let $\gamma_{a,D,X}$ to be dependent on the BS location $X$ so that every BS can make their own choice of UL/DL time split in an access subframe. We focus on $\gamma_{a,D,X} = \mathbb{1}(N_{d,X} > 0)\frac{N_{d,X}}{N_{u,X}+N_{d,X}}$, where $N_{u,X}$ and $N_{d,X}$ are the number of UL and DL users connected to the BS at $X$. Several variations of this policy are possible, such as adding a different optimized exponent $n$ to $N_{u,X}, N_{d,X}$ or incorporating other network parameters to capture the disparity of the UL/DL service rate. These variations are left to future work.

#### 4.3.2.2 Scheduling in backhaul subframes

Like the access subframe, it is possible to have static and dynamic TDD schemes for deciding the fraction of DL slots in a backhaul subframe. However, for analytical simplicity we assume $\gamma_{b,w,X} = \eta$, which is fixed for all $X \in \Phi_b$. Hierarchical scheduling is assumed in the backhaul subframe. First the MBSs make a decision of scheduling available SBSs with at least one UL/DL UE in a UL/DL backhaul subframe with uniformly random SBS selection for each slot. A SBS has to adhere to the slots allocated by its serving




(a) A TDD Frame.




(b) Figure shows (i) Heirarchical scheduling in backhaul subframe with UAB or SAB. (ii) Dynamic TDD can lead to different DL subframe sizes in access subframe.

Figure 4.2: TDD frame structure.

MBS for backhauling. Let the set $\mathcal{F}$ represent sub-frame lengths that are fixed across all BSs irrespective of the scheduling strategies. $F_a$ and $F_{bd}$ are two permanent members of $\mathcal{F}$. Further, $F_{ad}$ is also an element of $\mathcal{F}$ under static TDD scheme. Although $F_{bd}$ is fixed, a version of dynamic TDD is employed through UAB.

- **Synchronized access-backhaul (SAB).** SBS remains silent in unscheduled backhaul slots.

- **Unsynchronized access-backhaul (UAB) or poaching.** SBS schedules an UL/DL access link in the unscheduled backhaul slots. We focus here on a simple policy wherein UL access *poaches* only UL backhaul slots and similarly for DL. We assume that the SBS schedules an UL UE independently with probability $p_{ul}$ in an unscheduled backhaul UL slot and stays silent otherwise. $p_{dl}$ is the probability of scheduling a DL UE in a backhaul DL slot.

**Remark 4.** *The analysis of in-band backhauling in this chapter follows for out-of-band backhauling as well. In this case, a fraction $\delta$ of total bandwidth is allocated to access.*

### 4.3.3 Received signal power model

The received signal at $X \in \Phi_b \cup \Phi_u$ from $Y \in \Phi_b \cup \Phi_u$ with $X \neq Y$ in the $i^{\text{th}}$ time slot of a typical TDD frame is given by

$$P_r(X,Y) = \mathrm{C}_0 \mathrm{P}_Y h_{i,X,Y} G_{i,X,Y} L(X,Y)^{-1},$$

where $\mathrm{C}_0$ is the reference distance omnidirectional path loss at 1 meter, $\mathrm{P}_Y$ is the transmit power and is equal to either $\mathrm{P}_m$, $\mathrm{P}_s$ or $\mathrm{P}_u$ depending on whether $Y \in \Phi_m$, $Y \in \Phi_s$ or $Y \in \Phi_u$. $h_{i,X,Y}$ is the small scale fading, $G_{i,X,Y}$ is the product of transmit and receive antenna gains and $L(X,Y) = ||X-Y||^{\alpha_{X,Y}}$ is path loss between $X$ and $Y$. Here, $\alpha_{X,Y}$ is $\alpha_l$ with probability $p_l(||X-Y||)$ and $\alpha_n$ otherwise. There are several models proposed for $p_l(d)$ to incorporate blockage effects [24,34,46,53]. The generalized LOS ball model proposed in [46]

and validated in [47, 67] is used in this work. As per this model, $p_l(d) = p_{\text{LOS}}$ if $d \leq D_{\text{LOS}}$ and $p_l(d) = 0$ otherwise. Let $p_n(d) = 1 - p_l(d)$.

Here, $h_{i,X,Y}$ are independent and identically distributed (i.i.d.) to an exponential random variable with unit mean for all $X, Y \in \Phi_b \cup \Phi_u$. However, $h_{i,X,Y}$ can be arbitrarily correlated across time slots $i$. If the access link under consideration is a desired signal link, $G_{i,X,Y} = G_t G_u$, where $G_{(.)}$ denotes main lobe gain and $t \in \{m, s\}$. Similarly, $G_{i,X,Y} = G_m G_s$ for the backhaul desired signal link. An interfering link has antenna gain distribution as follows [34],

$$G_{i,X,Y} \overset{d}{=} \begin{cases} \Psi_{t_1,t_2} & \text{if } X \in \Phi_{t_1}, Y \in \Phi_{t_2} \\ & \text{with } t_1, t_2 \in \{m, s, u\} \text{ and } t_1 \neq t_2, \\ \Psi_{t,t} & \text{if } X, Y \in \Phi_t \text{ with } t \in \{m, s, u\}, \end{cases}$$

where $\overset{d}{=}$ denotes equality in distribution. Further, $G_{i,X,Y}$ is independently distributed with $G_{i,X',Y'}$ if at least one of $X \neq X'$ or $Y \neq Y'$. Also these gains are independent of $h_{i,X,Y}$, $\forall X, Y \in \Phi_b \cup \Phi_u$. Here, the probability mass functions (PMF) of $\Psi_{t,t}$ and $\Psi_{t_1,t_2}$ are given in Table 4.2, $g_{(.)}$ and $\Delta_{(.)}$ represent the side-lobe gain and 3-dB beam width.

Note that the correction factor introduced in Chapter 3 is not used in this work since the correction factor was discovered by the author after this study was performed. Also, since the focus here is on mean rate analysis (not cell edge rate analysis) in dense deployments, wherein there is a high probability of connecting with a LOS base station, the impact of correction factor on the rate performance will be negligible.

Table 4.2: Antenna gain distributions

| Parameter | Value | Probability |
|---|---|---|
| $\Psi_{t_1,t_2}$ | $G_{t_1}G_{t_2}$ | $\frac{\Delta_{t_1}\Delta_{t_2}}{4\pi^2}$ |
| | $G_{t_1}g_{t_2}$ | $\frac{\Delta_{t_1}(2\pi-\Delta_{t_2})}{4\pi^2}$ |
| | $g_{t_1}G_{t_2}$ | $\frac{(2\pi-\Delta_{t_1})\Delta_{t_2}}{4\pi^2}$ |
| | $g_{t_1}g_{t_2}$ | $\frac{(2\pi-\Delta_{t_1})(2\pi-\Delta_{t_2})}{4\pi^2}$ |
| $\Psi_{t,t}$ | $G_t^2$ | $\frac{\Delta_t^2}{4\pi^2}$ |
| | $G_tg_t$ | $\frac{2\Delta_t(2\pi-\Delta_t)}{4\pi^2}$ |
| | $g_t^2$ | $\frac{(2\pi-\Delta_t)^2}{4\pi^2}$ |

### 4.3.4 User and SBS association

Each user associates with either an MBS or SBS. Each SBS connects to an MBS. A typical user at $Z \in \Phi_u$ associates to BS at $X^*(Z) \in \Phi_b$ iff $X^*(Z) = \arg\max_{Y\in\Phi_t, t\in\{m,s\}} \mathrm{P}_t L(Y,Z)^{-1} G_t B_t$, where $B_t$ denotes a bias value multiplied to the received signal power from a BS of tier $t \in \{m, s\}$. Since the association criterion maps every point in $\Phi_u$ to a unique point in $\Phi_b$ almost surely, the mean number of users connected to a typical MBS is $\lambda_u\mathcal{A}_m/\lambda_m$, and that to a typical SBS is $\lambda_u\mathcal{A}_s/\lambda_s$ [46, 73]. Here, $\mathcal{A}_m$ is the probability of associating with a MBS and $\mathcal{A}_s = 1-\mathcal{A}_m$. The derivation of $\mathcal{A}_m$ can be found in Appendix 4.8.1. A SBS at $Z \in \Phi_s$ connects to a MBS at $X^*(Z) \in \Phi_m$ iff $X^*(Z) = \arg\min_{Y\in\Phi_m} L(Y,Z)$. Thus, the mean number of SBSs connected to a typical MBS is $\lambda_s/\lambda_m$.

### 4.3.5 Load distribution

Characterizing the load distribution with PPP BSs and UEs even under the simplest setting of nearest BS association is a long-standing open problem [74]. Several papers have assumed an independent load model for tractability [46, 73, 102–104]. Using a similar model, every $X \in \Phi_m$ is associated with independent marks $N_{s,X}, N_{u,X}, N_{d,X}$ representing number of SBSs, UL UEs and DL UEs connected to the MBS. Similarly, every $X \in \Phi_s$ is associated with independent marks $N_{u,X}, N_{d,X}$. Their distributional assumptions are given as follows [73, 103].

**Assumption 1.** *Let $\epsilon$ be the mean number of devices (users or SBSs) connected to a typical BS in $\Phi_t \in \{\Phi_m, \Phi_s\}$. The marginal probability mass function (PMF) of number of devices connected to a tagged and typical BS in $\Phi_b$ is given by $\kappa^*(n)$ and $\kappa(n)$ respectively.*

$$\kappa^*(n) = \frac{3.5^{3.5}\Gamma(n+3.5)\epsilon^{n-1}\left(3.5+\epsilon\right)^{-n-3.5}}{(n-1)!\Gamma(3.5)}, \text{ for } n \geq 1 \tag{4.1}$$

$$\kappa(n) = \frac{3.5^{3.5}\Gamma(n+3.5)\epsilon^{n}\left(3.5+\epsilon\right)^{-n-3.5}}{n!\Gamma(3.5)}, \text{ for } n \geq 0. \tag{4.2}$$

*Thus, the marginal PMF of $N_{s,X}, N_{u,X}, N_{d,X}$ is denoted as $\kappa_{s,t}, \kappa_{u,t}, \kappa_{d,t}$ for typical BS $X \in \Phi_t$ and with a superscript $*$ for tagged BS $X$. $\epsilon$ for each of these is given by $\frac{\lambda_s}{\lambda_m}$, $\frac{(1-\eta)\lambda_u \mathcal{A}_t}{\lambda_t}$ and $\frac{\eta\lambda_u \mathcal{A}_t}{\lambda_t}$, respectively.*

**Assumption 2.** *Let $\epsilon = \lambda_u \mathcal{A}_t/\lambda_t$ be the mean number of users connected to a typical BS in $\Phi_t \in \{\Phi_m, \Phi_s\}$. The joint PMF of number of UL and DL users*

*connected to a typical BS in* $\Phi_t$ *is given by* $\Upsilon_t(n_1, n_2, 3.5)$ *for* $n_1, n_2 \geq 0$*, where*

$$\Upsilon_t(n_1, n_2, k) = \frac{3.5^{3.5}}{\Gamma(3.5)} \frac{\eta^{n_2}(1-\eta)^{n_1}}{n_1! n_2!} \frac{\Gamma(n_1+n_2+k)}{\epsilon^k \left(1+\frac{3.5}{\epsilon}\right)^{n_1+n_2+k}}.$$

*Consider a BS serving the user at origin, then the joint PMF of number of UL and DL users connected to the BS* apart from the user at origin *is given by* $\Upsilon_t(n_1, n_2, 4.5)$ *for* $n_1, n_2 \geq 0$*.*

A summary of key notation is given in Table 4.1 and Fig. 4.2.

## 4.4 Uplink SINR and rate

As shown by Fig. 4.2, the SINR distribution will be dependent on the time slot $1 \leq i \leq \mathrm{F}$ and the scheduling strategies. Our goal is to compute the mean end-to-end rate of a typical user (UL or DL) at the origin under the various scheduling strategies described before. We analyze the marginal SINR distribution for access and backhaul links as two separate cases. Before going into the details, we first characterize the PMF of the number of DL access slots as follows.

**Lemma 8.** *The PMF of* $\mathrm{F}_{ad,w,X} \overset{d}{=} \mathrm{F}_{ad,w}$*, for a typical* $X \in \Phi_t$ *given* $\mathcal{F}$ *is computed as follows.*

*1. For static TDD, that is* $w = S$*,*

$$\mathbb{P}\left(\mathrm{F}_{ad,S,X} = n \middle| \mathrm{F}_a\right) = \tilde{\mathrm{F}}_{ad} \mathbb{1}\left(\lceil \gamma_a \mathrm{F}_a \rceil = n\right) + (1 - \tilde{\mathrm{F}}_{ad}) \mathbb{1}\left(\lfloor \gamma_a \mathrm{F}_a \rfloor = n\right), \quad (4.3)$$

*where* $\tilde{\mathrm{F}}_{ad} = \gamma_a \mathrm{F}_a - \lfloor \gamma_a \mathrm{F}_a \rfloor$*.*

*2. For dynamic TDD, that is $w = D$,*

$$\mathbb{P}\left(\mathrm{F}_{ad,D,X} = n \middle| \mathrm{F}_a\right) = \int_0^1 (p_1(n + r - 1) - p_2(n + 1 - r))\mathrm{d}r, \tag{4.4}$$

*where*

$$p_1(r) = \mathbb{1}(r > 0) \sum_{n_2=1}^{\infty} \sum_{n_1=0}^{\lceil \frac{n_2(\mathrm{F}_a - r)}{r} \rceil - 1} \Upsilon_t(n_1, n_2, 3.5) + \mathbb{1}(r \leq 0) - \mathbb{1}(r = 0) \left(1 + \frac{\mathcal{A}_t \lambda_u \eta}{3.5 \lambda_t}\right)^{-3.5},$$

$$p_2(r) = \mathbb{1}(r > 0) \sum_{n_2=1}^{\infty} \sum_{n_1=0}^{\lfloor \frac{n_2(\mathrm{F}_a - r)}{r} \rfloor} \Upsilon_t(n_1, n_2, 3.5) + \mathbb{1}(r \leq 0).$$

*Proof.* See Appendix 4.8.2. □

Small tail probabilities of the PMFs in Assumptions 1 and 2 for load values larger than the $\sim 6\times$ the mean allows us to compute the infinite sums as finite sums with first $\lfloor \frac{6\mathcal{A}_t \lambda_u}{\lambda_t} \rfloor$ terms.

### 4.4.1 SINR model for access links

Access links can be active in both access and backhaul subframes if the BSs operate in UAB. The SINR of a receiving BS at $X^* \in \Phi_t$, where $t \in \{m, s\}$, serving the UL user at origin is given as

$$\mathsf{SINR}^{ul}_{i,a,w} = \frac{\mathrm{C}_0 \mathrm{P}_u h_{i,X^*,0} G_u G_t L(X^*, 0)^{-1}}{I_{i,m,w}(X^*) + I_{i,s,w}(X^*) + I_{i,u,w}(X^*) + \sigma^2},$$

where $w \in \{S, D\}$ denotes static and dynamic TDD if $i \leq \mathrm{F}_a$ and $w \in \{\mathrm{SAB}, \mathrm{UAB}\}$ if $i > \mathrm{F}_a$. $I_{i,\nu,w}(Z)$ is the interference power at location $Z \in$

$\Phi_b \cup \Phi_u$ from all active devices of type $\nu \in \{m, s, u\}$ in the $i^{\text{th}}$ slot and $\sigma^2$ is the noise power. Here, for $\nu \in \{m, s\}$ and $i \leq \mathrm{F}_a$

$$I_{i,\nu,w}(Z) = \sum_{Y \in \Phi_\nu \backslash \{X^*\}} \mathbb{1}(i \leq \mathrm{F}_{ad,w,Y}) \mathbb{1}(N_{d,Y} > 0) \mathrm{C}_0 \mathrm{P}_\nu \times h_{i,Z,Y} G_{i,Z,Y} L(Z,Y)^{-1}. \quad (4.5)$$

Note that $\Phi_\nu \backslash \{X^*\} = \Phi_\nu$ if $X^* \notin \Phi_\nu$. Similarly, for $i \leq \mathrm{F}_a$

$$I_{i,u,w}(Z) = \sum_{Y \in \Phi_b \backslash \{X^*\}} \mathbb{1}(\mathrm{F}_{ad,w,Y} < i \leq \mathrm{F}_a) \mathbb{1}(N_{u,Y} > 0) \times \mathrm{C}_0 \mathrm{P}_u h_{i,Z,Y'} G_{i,Z,Y'} L(Z,Y')^{-1}, \quad (4.6)$$

where $Y'$ is the UL UE scheduled by BS at $Y$. If $i > \mathrm{F}_a$, then

$$I_{i,m,w}(Z) = \sum_{Y \in \Phi_m \backslash \{X^{**}\}} \mathbb{1}(\mathrm{F}_a < i \leq \mathrm{F}_a + \mathrm{F}_{bd}) \times \mathbb{1}\left(N_{s,d,Y} > 0\right) \mathrm{C}_0 \mathrm{P}_m h_{i,Z,Y} G_{i,Z,Y} L(Z,Y)^{-1}, \quad (4.7)$$

where $X^{**}$ is the location of MBS serving $X^* \in \Phi_s$ and $N_{s,d,Y}$ is the number of SBS with atleast one DL UE. Similarly, if $N_{s,u,Y}$ is the number of SBS connected to $Y \in \Phi_m$ with at least one UL UE,

$$\begin{aligned} I_{i,s,w}(Z) = \sum_{Y \in \Phi_m} & \mathbb{1}(\mathrm{F}_a + \mathrm{F}_{bd} < i \leq \mathrm{F}) \mathbb{1}\left(N_{s,u,Y} > 0\right) \\ & \times \mathrm{C}_0 \mathrm{P}_s h_{i,Z,Y'} G_{i,Z,Y'} L(Z,Y')^{-1} + \mathbb{1}(w = \mathrm{UAB}) \\ & \times \sum_{Y \in \Phi_s \backslash \{X^*\}} \mathbb{1}(\mathrm{F}_a < i \leq \mathrm{F}_a + \mathrm{F}_{bd}) \mathbb{1}\left(N_{d,Y} > 0\right) \\ & \times \xi_Y \zeta_Y \mathrm{C}_0 \mathrm{P}_s h_{i,Z,Y} G_{i,Z,Y} L(Z,Y)^{-1}, \quad (4.8) \end{aligned}$$

where $Y'$ is the SBS scheduled by MBS at $Y \in \Phi_m$. Here, $\zeta_Y$ is a Bernoulli random variable (independent across all $Y$) with success probability

$$p_{dl}\mathbb{1}\left(\mathrm{F}_a < i \leq \mathrm{F}_a + \mathrm{F}_{bd}\right) + p_{ul}\mathbb{1}\left(\mathrm{F}_a + \mathrm{F}_{bd} < i \leq \mathrm{F}\right),$$

and $\xi_Y$ is also an indicator random variable denoting whether the SBS is not scheduled by its serving MBS for backhauling in slot $i$ of the typical frame under consideration. Also,

$$I_{i,u,w}(Z) = \mathbb{1}(w = \text{UAB}) \sum_{Y \in \Phi_s \setminus \{X^*\}} \mathbb{1}(\mathrm{F}_a + \mathrm{F}_{bd} < i \leq \mathrm{F}) \\ \mathbb{1}(N_{u,Y} > 0)\xi_Y \zeta_Y \mathrm{C}_0 \mathrm{P}_u h_{i,Z,Y'} G_{i,Z,Y'} L(Z, Y')^{-1}, \quad (4.9)$$

where $Y' \in \Phi_{ul}$ is the UL user scheduled by the BS at $Y$.

(4.5)-(4.9) are applicable for evaluating the UL backhaul, DL access, and DL backhaul SINR distribution as well, although the receiving location $Z$ will be different under each case and is summarized in Table 4.3. Note that an UL access link will be active in a backhaul subframe only in $\mathrm{F}_a + \mathrm{F}_{bd} \leq i \leq \mathrm{F}$ and $w =$UAB scenario. Thus, to compute UL access SINR, (4.8) would have only the first summation term, and (4.7) would be zero.

**Remark 5** (A note on the interfering point processes in (4.5) to (4.9))**.** *Computing the Laplace transform of interference is a key step in evaluating SINR distribution. Exact expressions are available in literature for interferers generated from a PPP, Poisson cluster process, some special repulsive point processes [60, 105, 106]. Note that* (4.5) *and* (4.7) *have PPP interferers, and*

Table 4.3: Transmitter-receiver pairs for computing end-to-end rate of a typical user at origin.

| **Link** | **Receiver** | **Transmitter** |
|---|---|---|
| UL access | $X^*$ | 0 |
| UL backhaul | $X^{**}$ | $X^*$ |
| DL access | 0 | $X^*$ |
| DL backhaul | $X^*$ | $X^{**}$ |

*thus computing exact Laplace transform is possible. However,* (4.6)*,* (4.8) *and* (4.9) *have non-Poisson interfering processes, for which it is highly non-trivial to characterize the Laplace transform. Several approximate PPP models have been proposed in literature for computing Laplace functional of the interfering point processes in* (4.6) *and first term in* (4.8)*, for example [107–110]. We follow a theme of PPP approximations for the same inspired from these works. To compute an approximate Laplace transform of* (4.9) *and second term in* (4.8) *we propose novel PPP approximations on the same lines as [107] and validate these approximations with Monte-Carlo simulations.*

### 4.4.2 SINR distribution for access links

**Definition 5.** Conditioned on $\mathcal{F}$, the SINR coverage of a typical UL access link is defined as $\mathsf{S}^{ul}_{i,a,w}(\tau) = \mathbb{P}\left(\mathsf{SINR}^{ul}_{i,a,w} > \tau \middle| \mathcal{F}\right)$, if slot $i \leq \mathrm{F}_a$ and $w \in \{S, D\}$. If $i > \mathrm{F}_a$, typical UL UE is scheduled only if $w =$ UAB and it connects to a SBS. Thus, the SINR coverage for $i > \mathrm{F}_a$ is given by $\mathsf{S}^{ul,t}_{i,a,w}(\tau) = \mathbb{P}\left(\mathsf{SINR}^{ul}_{i,a,w} > \tau \middle| X^* \in \Phi_t, \mathcal{F}\right)$ for $t = s$.

**Definition 6.** The Laplace transform of the interference at a typical UL access

$$L_\nu = \prod_{\mu_1,\mu_2\in\{l,n\}} \exp\left(-\int_{\left(\frac{R^{\alpha_\mu}\mathrm{P}_\nu G_\nu B_\nu}{\mathrm{P}_t G_t B_t}\right)^{1/\alpha_{\mu_1}}}^{\infty} \int_0^{2\pi} \mathbb{E}\left[\frac{p_{i,D,\nu}\hat{\lambda}_{\nu,\mu_1,\mu_2}(r,\theta)r}{1+\frac{(r^2+R^2-2rR\cos(\theta))^{\alpha_{\mu_2}/2}}{\mathsf{s}\mathrm{C}_0\mathrm{P}_\nu\Psi_{t,\nu}}}\right]\mathrm{d}r\mathrm{d}\theta\right). \tag{4.10}$$

---

receiver at $X^*$ conditioned on the event that the receiving BS is at a distance $R$ and belongs to $\Phi_{t,\mu}$, which is the point process of LOS/NLOS BSs in $\Phi_t$ looking from origin, is given as follows for $\mu \in \{l, n\}$.

$$L_{i,w}^{ul,a,t,\mu}(\mathsf{s}, R) = \mathbb{E}\left[\exp\left(-\mathsf{s}I\right)\Big| X^* \in \Phi_{t,\mu}, ||X^*|| = R, \mathcal{F}\right],$$

where $I = I_{i,m,w}(X^*) + I_{i,s,w}(X^*) + I_{i,u,w}(X^*)$.

**Lemma 9.** *For $i \leq \mathrm{F}_a$, the Laplace transform $L_{i,w}^{ul,a,t,\mu}(\mathsf{s}, R) \approx L_m L_s L_u$, where*

- *For $\nu \in \{m, s\}$, $L_\nu = 1$ if $w = S$ and is given as follows if $w = D$,*

$$L_\nu \geq \exp\left(-\int_0^\infty \mathbb{E}\left[\frac{1}{1+\frac{r}{\mathsf{s}\mathrm{C}_0\mathrm{P}_\nu\Psi_{t,\nu}}}\right] p_{i,D,\nu}\Lambda_\nu(\mathrm{d}r)\right).$$

  *Exact expression for $L_\nu$ is given in* (4.10)*. where $\hat{\lambda}_{\nu,\mu_1,\mu_2}(r,\theta)$ is equal to*

$$\lambda_\nu p_{\mu_1}(r) p_{\mu_2}\left(\sqrt{r^2+R^2-2rR\cos\theta}\right),$$

  $p_{i,D,\nu} = \sum_{n=i}^{\mathrm{F}_a} \mathbb{P}\left(\mathrm{F}_{ad,D} = n\middle|\mathcal{F}\right)$*, and $\Lambda_\nu(\mathrm{d}\tau)$ is given in* (4.15)*. The expectation is with respect to the antenna gains $\Psi_{(.)}$ given in Table 4.2.*

- *For $w \in \{S, D\}$,*

$$L_u = \exp\left(-\int_0^\infty \mathbb{E}\left[\frac{1}{1+\frac{r}{\mathsf{s}\mathrm{C}_0\mathrm{P}_u\Psi_{t,u}}}\right]\Lambda(t,\mathrm{d}r)\right),$$

*where the expectation is with respect to the antenna gains* $\Psi_{(.)}$ *given in Table 4.2,* $\Lambda(t, \mathrm{d}r) = \sum_{k\in\{m,s\}} p_{i,w,k} \times$

$$\left(1 - \exp\left(-\Lambda_k\left(r\frac{\mathrm{P}_k B_k G_k}{\mathrm{P}_t B_t G_t}\right)\right)\right)\Lambda_k(\mathrm{d}r),$$

*with* $\Lambda_k(r)$ *given in* (4.14).

$$p_{i,S,k} = \left(1 - \left(1 + \frac{\lambda_u \mathcal{A}_k(1-\eta)}{3.5\lambda_k}\right)^{-3.5}\right)\mathbb{1}\left(\mathrm{F}_{ad} < i \leq \mathrm{F}_a\right),$$

*and*

$$p_{i,D,k} = \mathbb{P}\left(\mathrm{F}_{ad,D} < i \leq \mathrm{F}_a \middle| \mathcal{F}\right) - \left(1 + \frac{\lambda_u \mathcal{A}_k(1-\eta)}{3.5\lambda_k}\right)^{-3.5},$$

*which is computed using distribution of* $\mathrm{F}_{ad,D}$ *given in Lemma 8.*

**Lemma 10.** *For* $i > \mathrm{F}_a$ *and* $w = \mathrm{UAB}$*, the Laplace transform* $L_{i,w}^{ul,a,t,\mu}(\mathsf{s}, R) \approx L_s L_u$*, where*

$$L_s = \exp\left(-\int_0^\infty \mathbb{E}\left[\frac{1}{1 + \frac{r}{\mathsf{s}\mathrm{C}_0\mathrm{P}_s\Psi_{t,s}}}\right] \times \right.$$
$$\left. \left(p_{void}\left(1 - \exp\left(-\Lambda_m(r)\right)\right) + \exp\left(-\Lambda_m(r)\right)\right)\Lambda_m(\mathrm{d}r)\right),$$

$$L_u = \exp\left(-\int_0^\infty \mathbb{E}\left[\frac{1}{1 + \frac{r}{\mathsf{s}\mathrm{C}_0\mathrm{P}_u\Psi_{t,u}}}\right]\frac{\hat{\lambda}}{\lambda_s}\left(1 - \exp\left(-\Lambda_s(r)\right)\right)\Lambda_s(\mathrm{d}r)\right).$$

*Here, the expectation is with respect to the antenna gains* $\Psi_{(.)}$ *given in Table 4.2,*

$$p_{void} = 1 - \left(1 + \frac{\lambda_{s,u}}{3.5\lambda_m}\right)^{-3.5},$$

*with*

$$\lambda_{s,u} = \lambda_s\left(1 - \left(1 + \frac{\mathcal{A}_s\lambda_u(1-\eta)}{3.5\lambda_s}\right)^{-3.5}\right),$$

$$\hat{\lambda} = p_{ul}\left(\lambda_s - \left(1 - \left(1 + \frac{\lambda_s}{3.5\lambda_m}\right)^{-3.5}\right)\lambda_m\right)^{+} \times \mathbb{1}\left(\mathrm{F}_a + \mathrm{F}_{bd} < i \leq \mathrm{F}\right)\left(1 - \left(1 + \frac{\lambda_u(1-\eta)\mathcal{A}_s}{3.5\lambda_s}\right)^{-3.5}\right).$$

*Proof.* See Appendix 4.8.3 for proofs of Lemma 9 and Lemma 10.

**Theorem 6.** *For $i \leq \mathrm{F}_a$, the SINR coverage of a typical UL user is given by $\mathbb{E}\left[\mathsf{S}^{ul}_{i,a,w}(\tau)\right]$ where the expectation is over $\mathcal{F}$. For $i > \mathrm{F}_a$ and $w = \mathrm{UAB}$, the SINR coverage is given by $\mathbb{E}\left[\mathsf{S}^{ul,s}_{i,a,\mathrm{UAB}}(\tau)\right]$. Here, $\mathsf{S}^{ul}_{i,a,w}(\tau) = \mathcal{A}_s\mathsf{S}^{ul,s}_{i,a,w}(\tau) + \mathcal{A}_s\mathsf{S}^{ul,s}_{i,a,w}(\tau)$, where $\mathsf{S}^{ul,t}_{i,a,w}(\tau) =$*

$$\sum_{\mu\in\{l,n\}}\int_0^\infty \exp\left(\frac{-\tau R^{\alpha_\mu}\sigma^2}{\mathrm{C}_0\mathrm{P}_u G_u G_t}\right) L^{ul,a,t,\mu}_{i,w}\left(\frac{\tau R^{\alpha_\mu}}{\mathrm{C}_0\mathrm{P}_u G_u G_t}, R\right) \times \prod_{\substack{t'\in\{m,s\},\\ \mu'\in\{l,n\},\\ t'\neq t \text{ or } \mu'\neq\mu}} F_{t',\mu'}\left(\left(\frac{P_{t'}G_{t'}B_{t'}R^{\alpha_\mu}}{P_t B_t G_t}\right)^{\frac{1}{\alpha_{\mu'}}}\right)\frac{f_{t,\mu}(R)}{\mathcal{A}_t}\mathrm{d}R, \quad (4.11)$$

*where $L^{ul,a,t,\mu}_{i,w}(.)$ is given in Lemma 9 and 10,*

$$\begin{aligned}
F_{t,n}(R) &= \exp\left(-\pi\lambda_t\left(R^2 - p_{\mathrm{LOS}}\min(R, D_{\mathrm{LOS}})^2\right)\right),\\
F_{t,l}(R) &= \exp\left(-\pi\lambda_t p_{\mathrm{LOS}}\min(R, D_{\mathrm{LOS}})^2\right),\\
f_{t,l}(R) &= 2\pi\lambda_t R p_{\mathrm{LOS}}\mathbb{1}(R \leq D_{\mathrm{LOS}})\\
&\quad\times\exp\left(-\pi\lambda_t p_{\mathrm{LOS}}\min(R, D_{\mathrm{LOS}})^2\right),\\
f_{t,n}(R) &= 2\pi\lambda_t R\left(1 - p_{\mathrm{LOS}}\mathbb{1}(R \leq D_{\mathrm{LOS}})\right)\\
&\quad\times\exp\left(-\pi\lambda_t\left(R^2 - p_{\mathrm{LOS}}\min\left(R, D_{\mathrm{LOS}}\right)^2\right)\right).
\end{aligned}$$

*Proof.* The $\mathsf{SINR}$ coverage of a typical UL user scheduled in the $i^{\text{th}}$ slot ($i \leq \mathrm{F}_a$), is given by [9]

$$\begin{aligned}
&\mathsf{S}^{ul}_{i,a,w}(\tau) = \mathbb{P}\left(\mathsf{SINR}^{ul}_{i,a,w} > \tau\right) \\
&= \sum_{t\in\{s,m\},\, \mu\in\{l,n\}} \mathbb{P}\left(\mathsf{SINR}^{ul}_{i,a,w} > \tau, X^* \in \Phi_{t,\mu}\right) \\
&= \sum_{t\in\{s,m\},\, \mu\in\{l,n\}} \int_0^\infty \mathbb{P}\left(\mathsf{SINR}^{ul}_{i,a,w} > \tau,\right. \\
&\left. X^* \in \Phi_{t,\mu} \middle| ||X^*_{t,\mu}|| = R\right) f_{t,\mu}(R)\mathrm{d}R,
\end{aligned}$$

where $f_{t,\mu}(R)$ is the probability that there exists $X^*_{t,\mu}$, which is the BS nearest to origin of tier $t$ and link type $\mu \in \{l, n\}$, and its distance from origin is $R$. It is given as

$$\begin{aligned}
f_{t,\mu}(R) &= -\frac{\mathrm{d}}{\mathrm{d}R}\mathbb{P}\left(\Phi_{t,\mu}\left(\mathcal{B}(0,R)=0\right), \Phi_{t,\mu}\left(\mathbb{R}^2 > 0\right)\right) \\
&= 2\pi\lambda_t R p_\mu(R) \exp\left(-2\pi\lambda_t \int_0^R p_\mu(r) r \mathrm{d}r\right).
\end{aligned}$$

The SINR coverage expression is simplified further as follows. $\mathsf{S}^{ul}_{i,a,w}(\tau) =$

$$\begin{aligned}
&\sum_{t\in\{s,m\},\, \mu\in\{l,n\}} \int_0^\infty \mathbb{P}\left(\mathsf{SINR}^{ul}_{i,a,w} > \tau \middle| X^* \in \Phi_{t,\mu}, ||X^*|| = R\right) \times \\
&\qquad\qquad \mathbb{P}\left(X^* \in \Phi_{t,\mu} \middle| ||X^*|| = R\right) f_{t,\mu}(R)\mathrm{d}R \\
&= \sum_{\substack{t\in\{s,m\},\\ \mu\in\{l,n\}}} \int_0^\infty \mathbb{E}\left[e^{\frac{-\tau R^{\alpha_\mu}\left(I_{i,m,w}(X^*)+I_{i,s,w}(X^*)+I_{i,u,w}(X^*)+\sigma^2\right)}{\mathrm{C}_0 \mathrm{P}_u G_u G_t}} \middle| X^* \in \Phi_{t,\mu}, ||X^*|| = R\right] \\
&\times \prod_{t'\in\{m,s\},\mu'\in\{l,n\},t'\neq t \text{ or } \mu'\neq\mu} F_{t',\mu'}\left(\left(\frac{P_{t'}G_{t'}B_{t'}R^{\alpha_\mu}}{P_t B_t G_t}\right)^{1/\alpha_{\mu'}}\right) f_{t,\mu}(R)\mathrm{d}R.
\end{aligned}$$

[9] Note that conditioning on $\mathcal{F}$ is not explicitly written in the following equations for convenience.

where $F_{t,\mu}(R) = \mathbb{P}\left(\Phi_{t,\mu}\left(\mathcal{B}(0,R)=0\right)\right)$ and $\mathcal{B}(0,R)$ is the ball of radius R centered at the origin. $\square$

### 4.4.3 SINR distribution for backhaul links

SINR model for backhaul links is given by

$$\mathsf{SINR}^{ul}_{i,b,w} = \frac{\mathrm{C}_0 \mathrm{P}_s h_{i,X^*,X^{**}} G_m G_s L(X^*,X^{**})^{-1}}{I_{i,s,w}(X^{**}) + I_{i,u,w}(X^{**}) + \sigma^2},$$

where $w \in \{\mathrm{UAB}, \mathrm{SAB}\}$, $\mathrm{F}_a + \mathrm{F}_{bd} < i \leq \mathrm{F}$, $I_{i,u,\mathrm{SAB}} = 0$. $I_{i,s,w}(.)$ and $I_{i,u,\mathrm{UAB}}(.)$ are same as (4.8) and (4.9), respectively, except that here the receiver is $X^{**}$, which is the MBS serving the tagged SBS.

For the backhaul links, we are interested to find $\mathbb{P}\left(\mathsf{SINR}^{ul}_{i,b,w} > \tau \middle| X^* \in \Phi_s\right)$ where the probability is under the Palm of the user process. The reason is that for computing the end-to-end rate of a typical user at origin, we are interested in the distribution of backhaul SINR distribution only in scenarios when the user at origin connects to a SBS. However, to compute even serving distance distribution of backhaul link under the Palm of user process is highly non-trivial. In [100], such distribution was computed in the case when there were no blockage effects. Although in principle, such computations can be done with blockage effects there will be a total 12 cases that will arise – condition LOS/NLOS links for typical UE at origin to $X^*$ and the backhaul links between $X^*$ and $X^{**}$, and 3 sub-cases for each of these that account for different exclusion regions as shown in [100]. Computing the SINR CCDF under the Palm of the SBS process is much easier as follows and we will approximate the

SINR coverage of a typical backhaul link to be equal to that of the tagged link for rate computations, as also done previously in [46,98,99]. Validation of this is done in Figure 4.5b. Similar to UL access, the following can be derived.

**Corollary 4.** *CCDF of a typical backhaul UL SINR link for $i > \mathrm{F}_a$ is given as*

$$\mathsf{S}^{ul}_{i,b,w}(\tau) = \sum_{\mu\in\{l,n\}} \int_0^\infty \exp\left(\frac{-\tau R^{\alpha_\mu}\sigma_n^2}{\mathrm{C}_0\mathrm{P}_s G_s G_m}\right) \times L^{ul,b}_{i,w}\left(\frac{\tau R^{\alpha_\mu}}{\mathrm{C}_0\mathrm{P}_s G_s G_m}\right) F_{m,\mu'}\left(R^{\alpha_\mu/\alpha_{\mu'}}\right) f_{m,\mu}(R)\mathrm{d}R$$

*where* $L^{ul,b}_{i,w}(\mathsf{s}) = \mathbb{E}\left[\exp\left(-\mathsf{s}(I_{i,s,w}(X^{**}) + I_{i,u,w}(X^{**}))\right)\right] \approx L_s L_u$ *with*

$$L_s = \exp\left(-\int_0^\infty \mathbb{E}\left[\frac{1}{1+\frac{r}{\mathsf{s}\mathrm{C}_0\mathrm{P}_s\Psi_{m,s}}}\right] \times \left(1-\left(1+\frac{\lambda_{s,u}}{3.5\lambda_m}\right)^{-3.5}\right)\left(1-\exp\left(-\Lambda_m(r)\right)\right)\Lambda_m(\mathrm{d}r)\right),$$

*where* $\lambda_{s,u} = \lambda_s\left(1-\left(1+\frac{\lambda_u(1-\eta)\mathcal{A}_s}{3.5\lambda_s}\right)^{-3.5}\right)$.

$$L_u = \mathbb{1}(w=\mathrm{SAB}) + \mathbb{1}(w=\mathrm{UAB})\times \exp\left(-\int_0^\infty \mathbb{E}\left[\frac{\left(1-\exp\left(-\Lambda_s(r)\right)\right)\frac{\bar{\lambda}_u}{\lambda_s}\Lambda_s(\mathrm{d}r)}{1+\frac{r}{\mathsf{s}\mathrm{C}_0\mathrm{P}_u\Psi_{m,u}}}\right]\right),$$

*with*

$$\bar{\lambda}_u = p_{ul}\left(1-\left(1+\frac{\lambda_u(1-\eta)\mathcal{A}_s}{3.5\lambda_s}\right)^{-3.5}\right)\times \left(\lambda_s - \left(1-\left(1+\frac{\lambda_s}{3.5\lambda_m}\right)^{-3.5}\right)\lambda_m\right)^+.$$

*The expectation in the expressions for $L_u$ and $L_s$ is with respect to the antenna gains $\Psi_{(.)}$ given in Table 4.2.*

### 4.4.4 Mean rate analysis

Let $\mathcal{E}_m$ and $\mathcal{E}_s$ denote the events when the typical UE connects to a MBS and SBS, respectively.

**Typical UE connected to MBS.** Data transmitted by a typical UL user in a frame is given by

$$\mathrm{D}_{ul,m,w_a} = \mathrm{WT} \sum_{i=1+\mathrm{F}_{ad}}^{\mathrm{F}_a} \mathbb{1}\left(\text{UE scheduled in } i^{\text{th}} \text{ slot}\right) \log_2\left(1+\mathsf{SINR}^{ul}_{i,a,w_a}\right).$$

Here, $w_a \in \{S, D\}$ representing static and dynamic TDD. As time progresses, in every frame the data transmitted by the UL UE is distributed according to the above equation. Thus, the data rate of the user averaged over time is given by $\mathbb{E}\left[\mathrm{D}_{ul,m,w_a}\middle|\Phi_b,\Phi_u,\mathcal{E}_m\right]/\mathrm{TF}$, where expectation is over temporally varying random variables (all the randomness except that from $\Phi_b$ and $\Phi_u$). Spatial averaging over the user and BS point processes gives data rate of the typical user at origin as $\mathrm{R}_{ul,m,w_a} = \frac{\mathbb{E}\left[\mathrm{D}_{ul,m,w_a}\middle|\mathcal{E}_m\right]}{\mathrm{TF}}$.

**Typical UE connected to SBS.** Data transmitted by a typical UL user in access and backhaul slots of a typical frame is given by $\mathrm{D}_{ul,s,a,w_a}$ and $\mathrm{D}_{ul,s,b,w_b}$, which are given as follows.

$$\mathrm{D}_{ul,s,a,w_a} = \mathrm{WT} \sum_{i=1+\mathrm{F}_{ad}}^{\mathrm{F}} \mathbb{1}\left(\text{UE scheduled in } i^{\text{th}} \text{ slot}\right) \log_2\left(1+\mathsf{SINR}^{ul}_{i,a,w_a}\right),$$

$$\mathrm{D}_{ul,s,b,w_b} = \mathrm{WT} \sum_{i=1+\mathrm{F}_a+\mathrm{F}_{bd}}^{\mathrm{F}} \log_2\left(1+\mathsf{SINR}^{ul}_{i,b,w_b}\right) \times \mathbb{1}\left(\text{tagged SBS scheduled in } i^{\text{th}} \text{ slot and tx the UE's data}\right).$$

Here, $w_a \in \{S, D\}$ for access links and $w_b \in \{\text{UAB}, \text{SAB}\}$ for backhaul links. The data rate of the UE averaging over temporally varying random variables is given by $\tilde{R} =$

$$\frac{\min\left(\mathbb{E}\left[\mathrm{D}_{ul,s,a,w_a}\middle|\Phi_b,\Phi_u,\mathcal{E}_s\right],\mathbb{E}\left[\mathrm{D}_{ul,s,b,w_b}\middle|\Phi_b,\Phi_u,\mathcal{E}_s\right]\right)}{\mathrm{TF}}.$$

The data rate after spatial averaging is given by expectation of the aforementioned rate over $\Phi_b$ and $\Phi_u$ and is given by $\mathrm{R}_{ul,s,w_a,w_b} = \mathbb{E}\left[\tilde{R}\middle|\mathcal{E}_s\right] \leq \min\left(\mathbb{E}\left[\mathrm{D}_{ul,s,a,w_a}\middle|\mathcal{E}_s\right],\mathbb{E}\left[\mathrm{D}_{ul,s,b,w_b}\middle|\mathcal{E}_s\right]\right)/\mathrm{TF}$. We will use this upper bound as an approximation to our mean rate estimates. We observe that the upper bound is very close to actual mean rate for $\delta$ larger or smaller than the optimal $\delta \in \{0.1, 0.2, \ldots, 0.9\}$, which is intuitive since the network is either highly access or backhaul limited in these scenarios and thus the minimum of expectation is roughly equal to expectation of minimum. For $\delta$ close to the optimal, there is some gap and in future it will be desirable to close this gap with a better approximation. However, the estimates for optimal $\delta$ were observed to be roughly same with the upper bound and the actual ergodic mean rate [111].

**Theorem 7.** *Approximate mean rate of a typical UL user in the network is given by* $\mathrm{R}_{ul,w_a,w_b} = \mathcal{A}_m \mathrm{R}_{ul,m,w_a} + \mathcal{A}_s \mathrm{R}_{ul,s,w_a,w_b}$, *where*

$$\mathrm{R}_{ul,m,w_a} = \mathbb{E}_{\mathcal{F}} \frac{\mathrm{W}}{\mathrm{F}} \sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty} \frac{\Upsilon_m(n_1,n_2,4.5)}{n_1+1} \int_0^{\infty} \frac{\sum_{i=1+\mathrm{F}_{ad,w_a,X^*}}^{\mathrm{F}_a} \mathsf{S}_{i,a,w_a}^{ul,m}(\tau)}{1+\tau} \mathrm{d}\tau.$$

$\mathbb{E}_{\mathcal{F}}$ *denotes expectation is over* $\mathcal{F}$. *Also note that given the UL and DL loads* $n_1$ *and* $n_2$, $\mathrm{F}_{ad,w_a,X^*}$ *is computed as per Section 4.3.2.*

$$\mathrm{R}_{ul,s,w_a,w_b} = \mathbb{E}_{\mathcal{F}} \frac{\min\left(\mathrm{R}_{a,ul,s,w_a,w_b}, \mathrm{R}_{b,ul,s,w_a,w_b}\right)}{\mathrm{F}},$$

$$\mathrm{R}_{a,ul,s,w_a,w_b} = \mathrm{W}\sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty}\frac{\Upsilon_s(n_1,n_2,4.5)}{n_1+1}\int_0^{\infty}\frac{\sum_{i=1+\mathrm{F}_{ad,w_a,X^*}}^{\mathrm{F}_a}\mathsf{S}_{i,a,w_a}^{ul,s}(\tau)}{1+\tau}\mathrm{d}\tau$$
$$+\mathbb{1}(w_b=\mathrm{UAB})\mathrm{W}(1-\mathbb{E}\left[1/N_{s,u}\right])p_{ul}\sum_{n=1}^{\infty}\frac{\kappa_{u,s}^*(n)}{n}\int_0^{\infty}\frac{\sum_{i=1+\mathrm{F}_a+\mathrm{F}_{bd}}^{\mathrm{F}}\mathsf{S}_{i,a,w_a}^{ul,s}(\tau)}{1+\tau}\mathrm{d}\tau, \tag{4.12}$$

---

*where* $\mathrm{R}_{a,ul,s,w_a,w_b}$ *is given in* (4.12) *and*

$$\mathrm{R}_{b,ul,s,w_a,w_b} = \mathrm{WT}\mathbb{E}\left[1/N_{s,u}\right]\sum_{n=1}^{\infty}\frac{\kappa_{u,s}^*(n)}{n}\int_0^{\infty}\frac{\sum_{i=1+\mathrm{F}_a+\mathrm{F}_{bd}}^{\mathrm{F}}\mathsf{S}_{i,b,w_b}^{ul}(\tau)}{1+\tau}\mathrm{d}\tau,$$

*where* $w_a \in \{S,D\}$, $w_b \in \{\mathrm{SAB},\mathrm{UAB}\}$, $N_{s,u}$ *has distribution in* (4.1) *with* $\epsilon = \lambda_s\left(1-\left(1+\frac{\mathcal{A}_s\lambda_u(1-\eta)}{3.5\lambda_s}\right)^{-3.5}\right)/\lambda_m$. *Also,* $\kappa_{u,s}^*$, $\Upsilon_m(.)$ *and* $\Upsilon_s(.)$ *are given in Section 4.3.5. Further, the notation* $\sum_x^y$ *implicitly assumes that the sum is zero if* $y<x$.

*Proof.* See Appendix 4.8.4. □

**Remark 6.** *The infinite summations in Theorem 7 correspond to averaging some load distribution, as inferred from Appendix 4.8.4. These can be computed accurately as finite sums with roughly* $6x$ *terms if the mean load for the particular summation is* $x$.

## 4.5 Downlink SINR and rate

Analyzing DL SNR distribution is very similar to UL, and the key difference lies in the interference distribution which results due to the receiver

position now being at the origin instead of $X^*$ or $X^{**}$ as in the UL case. This leads to different exclusion regions that need to be considered while computing shot noise of the interfering points as will be clear in Appendix 4.8.5. For rate computations, another major difference arises due to different probability of being scheduled in $i^{\text{th}}$ slot for DL and UL UEs, that depends on the DL subframe length distribution in access and backhaul subframes as a function of $\eta$.

**SINR distribution for access links.** DL SINR of a typical UE at the origin being served by a BS at $X^* \in \Phi_t$, where $t \in \{m, s\}$, is given as follows.

$$\mathsf{SINR}^{dl}_{i,a,w} = \frac{\mathrm{C}_0 \mathrm{P}_t h_{i,0,X^*} G_u G_t L(0, X^*)^{-1}}{I_{i,m,w}(0) + I_{i,s,w}(0) + I_{i,u,w}(0) + \sigma^2},$$

where $w \in \{S, D\}$ if $i \leq \mathrm{F}_a$ and $w \in \{\mathrm{SAB}, \mathrm{UAB}\}$ if $i > \mathrm{F}_a$. $I_{i,\nu,w}(0)$ is the interference power at origin from all active devices of type $\nu \in \{m, s, u\}$ in the $i^{\text{th}}$ slot as given in (4.5)-(4.9). Note that here the DL access link will be active only when $\mathrm{F}_a < i \leq \mathrm{F}_a + \mathrm{F}_{bd}$ in the backhaul subframe and thus, the second sum in (4.8) would be non-zero but the first summation would be zero.

The SINR distribution is given similar to (4.11) and is given as follows,

$$\mathsf{S}^{dl,t}_{i,a,w}(\tau) =$$

$$\sum_{\substack{t\in\{s,m\},\\ \mu\in\{l,n\}}} \int_0^\infty \exp\left(\frac{-\tau R^{\alpha_\mu}\sigma^2}{\mathrm{C}_0 \mathrm{P}_t G_u G_t}\right) L^{dl,a,t,\mu}_{i,w}\left(\frac{\tau R^{\alpha_\mu}}{\mathrm{C}_0 \mathrm{P}_t G_u G_t}, R\right)$$

$$\times \prod_{\substack{t'\in\{m,s\},\\ \mu'\in\{l,n\},\\ t'\neq t \text{ or } \mu'\neq\mu}} F_{t',\mu'}\left(\left(\frac{P_{t'} G_{t'} B_{t'} R^{\alpha_\mu}}{P_t B_t G_t}\right)^{\frac{1}{\alpha_{\mu'}}}\right) \frac{f_{t,\mu}(R)}{\mathcal{A}_t} \mathrm{d}R. \quad (4.13)$$

Note the different transmit power here and also that $L_{i,w}^{dl,a,t,\mu}(\mathsf{s},R)$, derived in Appendix 4.8.5, is different from the UL Laplace transform of interference given in Lemmas 9 and 10.

**SINR distribution for backhaul links.** For DL backhaul link, considering a typical SBS at origin being served by a MBS at $X^{**}$,

$$\mathsf{SINR}_{i,b,w}^{dl} = \frac{\mathrm{C}_0 \mathrm{P}_m h_{i,X^{**},0} G_m G_s L(X^{**},0)^{-1}}{I_{i,s,w}(0) + I_{i,m,w}(0) + \sigma^2},$$

where $w \in \{\mathrm{UAB}, \mathrm{SAB}\}$, $\mathrm{F}_a < i \leq \mathrm{F}_a + \mathrm{F}_{bd}$, $I_{i,s,\mathrm{SAB}} = 0$. $I_{i,m,w}(0)$ and $I_{i,s,\mathrm{UAB}}(0)$ can be obtained from (4.7) and (4.8), respectively. $\mathsf{S}_{i,b,w}^{dl}$ is same as Corollary 4 with $L_{i,w}^{ul,b}$ replaced by $L_{i,w}^{dl,b}(\mathrm{s},\rho) \approx L_m L_s$, where

$$L_s = \exp\left(-\int_0^\infty \mathbb{E}\left[\frac{\bar{\lambda}_d \Lambda_s(\mathrm{d}r)/\lambda_s}{1+\frac{r}{\mathsf{s}\mathrm{C}_0\mathrm{P}_u\Psi_{m,u}}}\right]\right),$$

if $w = \mathrm{UAB}$ and $L_s = 1$ if $w = \mathrm{SAB}$. Here,

$$\bar{\lambda}_d = \left(\lambda_s - \left(1 - \left(1 + \frac{\lambda_s}{3.5\lambda_m}\right)^{-3.5}\right)\lambda_m\right)^+ \times \left(1 - \left(1 + \frac{\lambda_u \eta \mathcal{A}_s}{3.5\lambda_s}\right)^{-3.5}\right) p_{dl}.$$

and $L_m = \exp(-\theta)$, where

$$\theta = \int_{\rho^{\alpha_l}}^\infty \mathbb{E}\left[\frac{\left(1 - \left(1 + \frac{\lambda_{s,d}}{3.5\lambda_m}\right)^{-3.5}\right)\Lambda_m(\mathrm{d}r)}{1+\frac{r}{\mathsf{s}\mathrm{C}_0\mathrm{P}_m\Psi_{m,s}}}\right],$$

and $\lambda_{s,d} = \lambda_s\left(1 - \left(1 + \frac{\lambda_u\eta\mathcal{A}_s}{3.5\lambda_s}\right)^{-3.5}\right)$. The expectation in the expression for $L_s$ and $\theta$ is with respect to the antenna gains $\Psi_{(.)}$ given in Table 4.2.

**Theorem 8.** *The mean rate of a typical DL user in the network is given by* $\mathrm{R}_{dl,w_a,w_b} = \mathcal{A}_m \mathrm{R}_{dl,m,w_a} + \mathcal{A}_s \mathrm{R}_{dl,s,w_a,w_b}$, *where*

$$\mathrm{R}_{dl,m,w_a} = \mathbb{E}_{\mathcal{F}} \frac{\mathrm{W}}{\mathrm{F}} \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \frac{\Upsilon_m(n_1, n_2, 4.5)}{n_2+1} \int_0^{\infty} \frac{\sum_{i=1}^{\mathrm{F}_{ad,w_a,X^*}} \mathsf{S}_{i,a,w_a}^{dl,m}(\tau)}{1+\tau} \mathrm{d}\tau,$$

$$\mathrm{R}_{dl,s,w_a,w_b} = \mathbb{E}_{\mathcal{F}} \frac{\min\left(\mathrm{R}_{a,dl,s,w_a,w_b}, \mathrm{R}_{b,dl,s,w_a,w_b}\right)}{\mathrm{TF}},$$

$$\mathrm{R}_{a,dl,s,w_a,w_b} = \mathrm{WT} \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \frac{\Upsilon_s(n_1, n_2, 4.5)}{n_2+1} \int_0^{\infty} \frac{\sum_{i=1}^{\mathrm{F}_{ad,w_a,X^*}} \mathsf{S}_{i,a,w_a}^{dl,s}(\tau)}{1+\tau} \mathrm{d}\tau$$

$$+ \mathbb{1}(w_b = \mathrm{UAB}) \mathrm{WT}(1 - \mathbb{E}\left[1/N_{s,d}\right]) p_{dl} \sum_{n=1}^{\infty} \frac{\kappa_{d,s}^*(n)}{n} \int_0^{\infty} \frac{\sum_{i=1+\mathrm{F}_a}^{\mathrm{F}_{bd}} \mathsf{S}_{i,a,w_a}^{dl,s}(\tau)}{1+\tau} \mathrm{d}\tau,$$

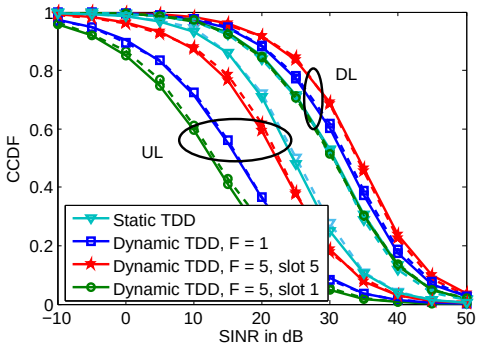
$$\mathrm{R}_{b,dl,s,w_a,w_b} = \mathrm{WT}\mathbb{E}\left[1/N_{s,d}\right] \sum_{n=1}^{\infty} \frac{\kappa_{d,s}^*(n)}{n} \int_0^{\infty} \frac{\sum_{i=1+\mathrm{F}_a}^{\mathrm{F}_{bd}} \mathsf{S}_{i,b,w_b}^{dl}(\tau)}{1+\tau} \mathrm{d}\tau,$$

*where* $w_a \in \{S, D\}$ *and* $w_b \in \{\mathrm{SAB}, \mathrm{UAB}\}$. *Here,* $N_{s,d}$ *has distribution as in* (4.1) *with* $\epsilon = \frac{\lambda_s \left(1 - \left(1 + \frac{\mathcal{A}_s \lambda_u \eta}{3.5 \lambda_s}\right)^{-3.5}\right)}{\lambda_m}$. *Also,* $\kappa_{d,s}^*$, $\Upsilon_m(.)$ *and* $\Upsilon_s(.)$ *are given in Section 4.3.5.*

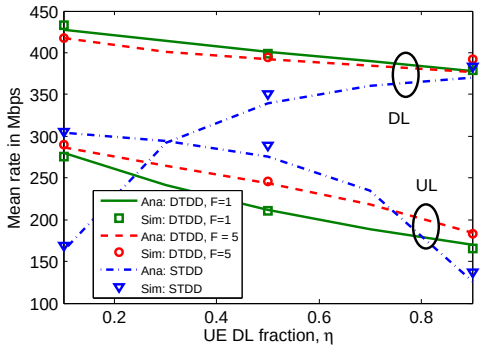
*Proof.* Follows Appendix 4.8.4. Note the subtle differences in the limits of summations inside the integrals here compared to Theorem 7. This is due to the different subframes in which an UL or DL UE or SBS is scheduled. □

**Corollary 5.** *The mean rate of a typical user is given by* $\mathrm{R}_{w_a,w_b} = \eta \mathrm{R}_{dl,w_a,w_b} + (1-\eta)\mathrm{R}_{ul,w_a,w_b}$.

*Proof.* The typical point at origin is DL with probability $\eta$ and UL with probability $1-\eta$. □




(a) SINR validation. Dotted lines are with Monte Carlo simulations.




(b) UL and DL Mean Rates

Figure 4.3: Impact of frame size and analysis validation.

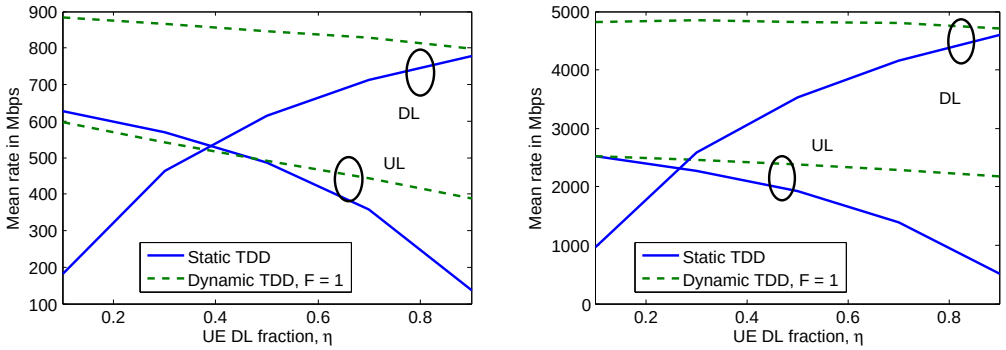
**Remark 7.** *We recommend to first evaluate the SINR coverage for different thresholds, and then use the stored values to compute the numerical integrals involved in the mean rate formulae. Our codes can be accessed at [111].*

## 4.6 Numerical results

First we study static vs dynamic TDD when all BSs are MBSs. Then we introduce wirelessly backhauled SBSs into the network and study the comparison of TDD schemes.

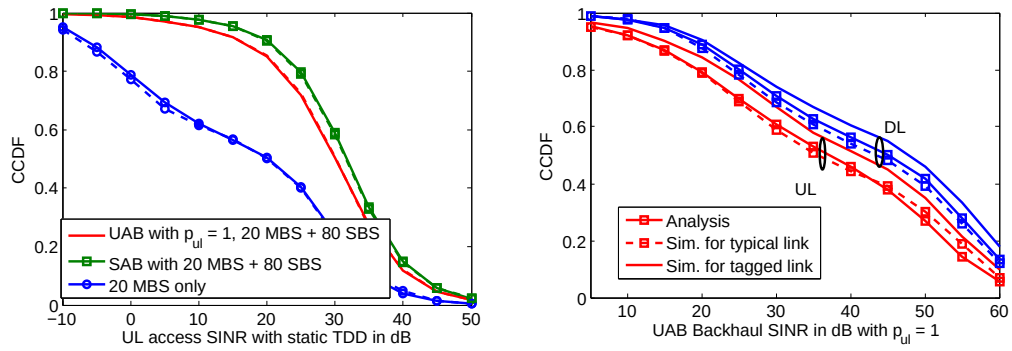
### 4.6.1 Dynamic vs static TDD when all BSs are MBSs

**Validation of analysis and impact of frame size.** Fig. 4.3(a) validates UL and DL SINR distribution with static and dynamic TDD for frame-size F = 1 and F = 5 with $\eta = 0.5$ and $\lambda_u = 500/\text{km}^2$. The Monte Carlo simulations match the analytical results very well. Fig. 4.3(a) also shows that




(a) $f_c = 28$ GHz, B = 200 MHz (b) $f_c = 73$ GHz, B = 2 GHz

Figure 4.4: Dynamic vs Static TDD, $\lambda_u = 200/\text{km}^2$. Dynamic TDD helps the "rare" UEs in the network perform much better.




(a) Uplink access. Dotted lines with Monte Carlo simulations. (b) Backhaul

Figure 4.5: SINR validation with self-backhauling for $\eta = 0.5$. Also shows self-backhauling is a good coverage solution.

UL SINR coverage with static TDD is better than with dynamic TDD but DL SINR coverage with dynamic TDD is better than with static TDD. This is primarily because of the transmit power disparity between UL and DL. For a moderately loaded system, as considered in this setup, the average number of interferers seen by a typical UL user is roughly the same with static and dynamic TDD. However, with dynamic TDD some of these interferers now have 10 dB more transmit power, which increases the interference and thus lowers SINR coverage. Note that the location of the interferers with static and dynamic TDD are different and thus a theoretical result like stochastic dominance of UL SINR with static TDD over dynamic TDD cannot be stated.

Fig. 4.3(a) further shows that the UL SINR coverage with dynamic TDD for slot 5 with F = 5 is better by about 10 dB than F = 1 and by 15 dB for slot 1 with F = 5, which is significant. This can be explained as follows. For F = 1, the probability that an interferer is DL is 0.5, whereas for F = 5 the probability rises to 0.95 (computed using the formula in Lemma 8) for slot 1 and decreases to 0.04 for slot 5. Since DL transmit power is much higher than UL, the UL SINR coverage for F = 1 falls between the two curves for F = 5. *Thus, there is an inherent UL interference mitigation with larger frame size* since UL UE has more chances on being scheduled towards end of the frame than at the beginning, as can be seen in Fig. 4.3(b). Similar observations can be made for DL but are less pronounced since DL to DL interference is less significant than DL to UL due to low UL transmit power.

**Dynamic TDD not desirable in high load interference-limited**

**scenarios but desirable in low load and asymmetric traffic scenarios.** Fig. 4.3(b) plots the UL and DL mean rates with static and dynamic TDD for different values of $\eta$. First, note that the analytical formula gives a close match with the Monte Carlo simulations. Dynamic TDD essentially helps boost the rates of the "rare" UEs in the network. For example, the DL rates double when $\eta = 0.1$ with dynamic TDD. In this scenario, there is about 5.6% loss in UL rate with dynamic TDD. Similarly, note the 1.5× gain for UL when $\eta = 0.9$. This indicates that dynamic TDD can be beneficial in asymmetric traffic scenarios but the gains are not very significant for $\eta$ close to 0.5, in fact there is 15% gain for DL but 11% loss for UL. *Thus, in high load interference-limited scenarios it is beneficial to switch to load aware static TDD.* The comparison is more persuasive for dynamic TDD in a low load scenario as shown in Fig. 4.4(a) and even more for noise-limited 73 GHz network with 2 GHz bandwidth as shown in Fig. 4.4(a). For example, Fig. 4.4(a) shows that the mean rates with DL (UL) are 5× with dynamic TDD for $\eta = 0.1(0.9)$. Even for $\eta = 0.5$, there is a gain of 23% for UL and 37% for DL. To summarize the observations for MBS only scenario: low load, asymmetric traffic, and noise-limitedness benefit dynamic TDD.

### 4.6.2 Impact of self-backhauling

**Validation of analysis.** Fig. 4.5 validates the SINR coverage for access and backhaul links for the 28GHz network under consideration, and a very close match is seen between analysis and Monte Carlo simulations. In

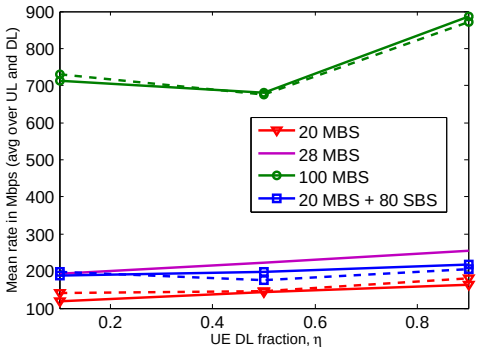
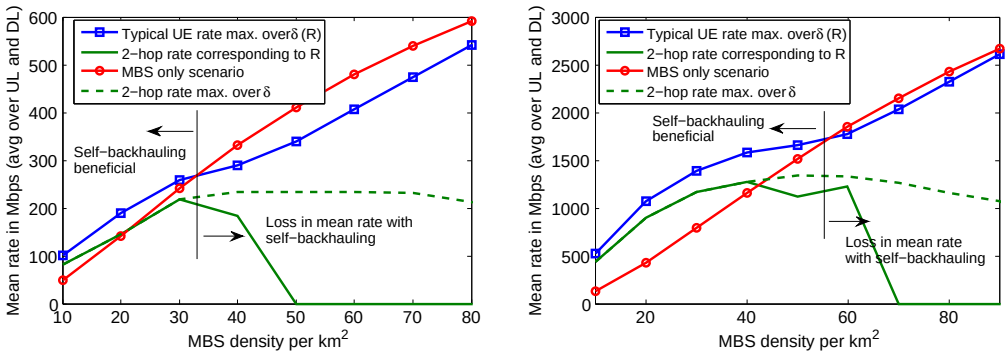



Figure 4.6: Self-backhauling is a poor substitute for wired backhauling. Dotted lines with Monte Carlo simulations.

Fig. 4.5(b) it can be seen that assuming typical SBS SINR instead of tagged SBS SINR can give an error of about 2-3dB, which is reasonable for analyzing mean rates as seen in Fig. 4.6.

**Low cost coverage solution but not for boosting mean rate.** Fig. 4.5(b) also shows that the $95^{\text{th}}$ percentile SINR increases by almost 20 dB when 80 additional SBSs are introduced to a baseline MBS only network. This clearly shows the coverage improvement with self-backhauling that translates into significant gain in cell edge rates. For example, here the cell edge rates go from $4.7 \times 10^6$ to $2.5 \times 10^7$ for $\eta = 0.5$. However, as can be seen from Fig. 4.6 the mean rates increase by only $33\% - 57\%$ across different $\eta$ after addition of 80 SBSs. This is equivalent to adding only 8 MBSs in terms of mean rate, although the 20 dB coverage improvement will not be seen in that case. Note that the mean rate values for the self-backhauling case in Fig. 4.6 are for static TDD with SAB and $\delta$ is chosen to be the maximizer of mean rates. If 80 MBSs were added instead of 80 SBSs, the rates increase by more

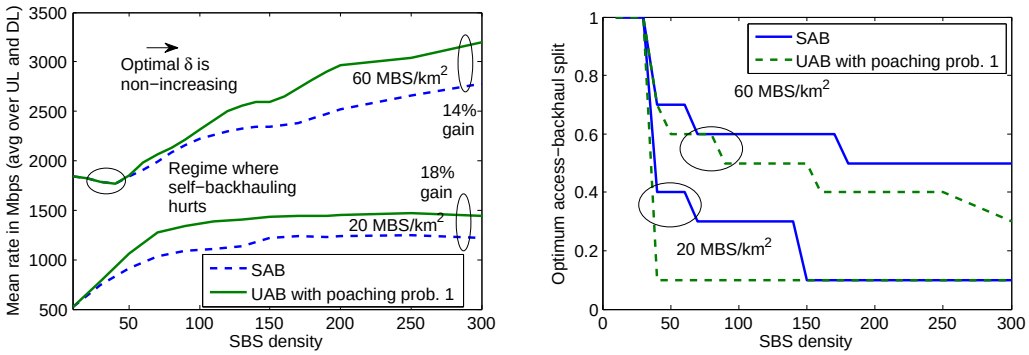



(a) 28 GHz, 200 MHz (b) 73 GHz, 2 GHz

Figure 4.7: Fix $\lambda_b = 100/\text{km}^2$ and vary $\lambda_m$. Optimization over $\delta$ is done by choosing the best from $\{0.1, 0.2, \ldots, 1\}$.

than 7× compared to baseline scenario. Thus, self-backhauling is a low cost coverage solution and not for increasing data rates.

**Trends with network densification.** Fig. 4.7 compares the mean rate of self-backhauled networks with $\lambda_b$ fixed at $100/\text{km}^2$ and varying $\lambda_m/\lambda_s$ and MBS only networks with $\lambda_m = 100/\text{km}^2$. *One would expect that adding SBSs on top of MBSs would always increase the rate. However, counter-intuitively this does not occur.* When MBS density is low, as expected adding SBSs such that total density is $100/\text{km}^2$ increases data rates. The rates shown in the Figure correspond to the access backhaul split that maximizes rate. When MBS density$\geq 50/\text{km}^2$ in Fig. 4.7(a) and $\geq 70/\text{km}^2$ in Fig. 4.7(b), the 2 hop rates corresponding to optimal $\delta$ go to zero implying $\delta = 1$. This occurs because the 2 hop rates are much lower than the single hop rates (the dotted line in the figure shows this wherein $\delta$ was chosen to maximize the 2 hop rate) and maximizing over mean rate kills the 2 hop rates to zero, giving



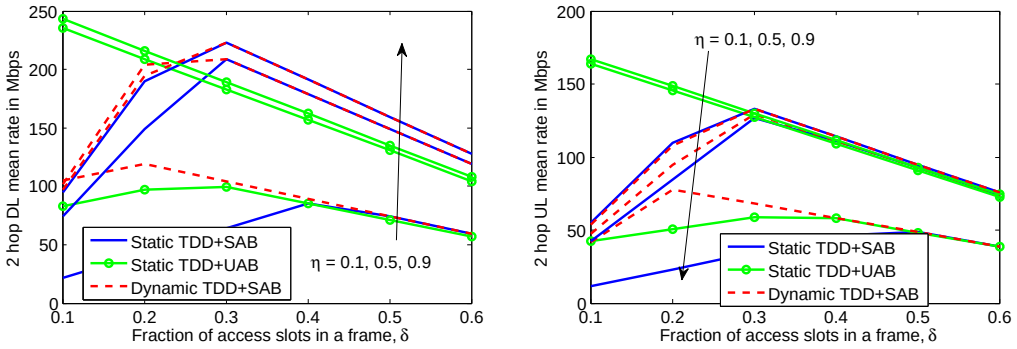

(a) SBS saturation density increases for higher MBS density.

(b) Network becomes backhaul limited with increasing SBS density.

Figure 4.8: Fix $\lambda_m$ and vary $\lambda_s$. Here, $\eta = 0.5$.

as many resources to direct links. This indicates that *when there are enough MBSs, adding just a few SBSs may not be beneficial as the slight benefit in coverage is overshadowed by the loss due to 2 hops.* The losses can be converted to no-loss by biasing UEs towards MBS. Fig. 4.7(b) corresponds to a noise-limited scenario and also in this case the DL access transmit power is reduced to 20dBm keeping backhaul transmit power as 30dBm as an example of a network which is less backhaul-limited. In this case the "beneficial" regime with self-backhauling is pushed further towards $\lambda_b$.

In Fig. 4.8(a), for a fixed $\lambda_m$, the value of $\lambda_s$ is increased. For each self-backhauling configuration an optimum $\delta$ is chosen from the set $\{0.1, 0.2, \ldots, 1\}$ and is shown in Fig. 4.8(b). *The optimum $\delta$ is non-increasing with SBS density and UAB as is expected.* Since more UEs connect with SBSs, we need more backhaul slots in a frame. There are another couple of observations to be made in Fig. 4.8(a). Firstly, note that UAB gives about $10 - 20\%$ gain over SAB.




(a) Downlink (b) Uplink

Figure 4.9: Comparison of TDD schemes across different $\delta$ and $\eta$, and impact on optimal $\delta$.

The gain is negligible or none at lower SBS densities wherein there are not many backhaul slots to be poached. Also note that *the rates saturate sooner in the 20 MBS case than the 60 MBS case*. As SBS density becomes large, the network becomes *backhaul limited* as indicated by the decreasing optimum $\delta$ in Fig. 4.8(b). Similar observations can be noted for the 28 GHz network, although the gains with UAB are negligible in that case due to increasing interference.

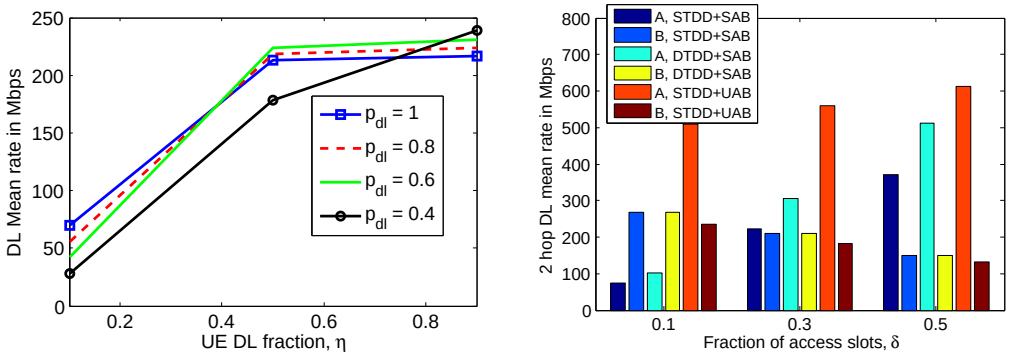
### 4.6.3 Comparison of TDD schemes

**Gains from Dynamic TDD and UAB held back by weak backhaul links.** Fig. 4.9 shows the comparison of 2 hop rates with different TDD schemes. As expected from our observations in Section 4.6.1, for $\eta = 0.1, 0.9$ dynamic TDD provides about 1.5× gains for DL/UL compared to static TDD for an optimal $\delta$ chosen for each scheme. For $\eta = 0.5$, the gains with dynamic

TDD are completely overshadowed by weak backhaul links for the optimum choice of $\delta$ but $20-30\%$ gains are visible for non-optimal $\delta$ lower than the optimal. Note that choosing a $\delta$ higher than optimum gives same rate as static TDD since the network is backhaul limited and this is the backhaul rate on the 2 hop link. This is clearer looking at the access and backhaul rates separately for DL UEs operating on 2 hops, as shown in Fig. 4.10(b). Another observation from Fig. 4.9 is that *the optimal $\delta$ with dynamic TDD and UAB is lower or the same as compared to static TDD with SAB*. The reason is that both dynamic TDD and UAB boost access rates for a fixed $\delta$ (see Fig. 4.10(b)) and thus can allow providing more backhaul slots in a frame still being able to achieve higher 2 hop rates. Fig. 4.10(b) also shows a potential of up to $2-5\times$ gains in DL rates with UAB for $\eta = 0.5$ and different $\delta$, but the gains are held back by weak backhaul links.

**UAB gains are not limited to asymmetric traffic.** Fig. 4.9 shows that with UAB, unlike dynamic TDD, about 30% gains are still observed in UL 2 hop rates for $\eta = 0.5$. The gains with DL are only 10% since due to increasing interference, $p_{dl} = 1$ is not optimal as seen from Fig. 4.10(a). Also, since DL access rates are closer to backhaul rates due to higher transmit power compared to UL, the network is even more backhaul-limited from DL UE perspective.

**Consistent 30% gains in mean rates across all traffic scenarios with dynamic TDD + UAB in a noise-limited scenario.** Finally, shifting our focus back to the 73 GHz network mentioned before, which had




(a) Optimal $p_{dl}$ is lower for higher $\eta$.

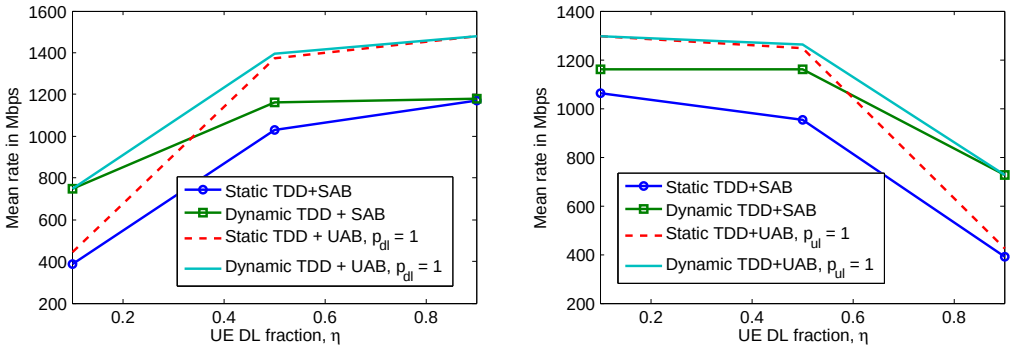
(b) 'A' for access and 'B' for backhaul. $p_{dl} = 1$, $\eta = 0.5$.

Figure 4.10: DL mean rates conditioned that UE connects to SBS.

stronger backhaul links, we can see in Fig. 4.11 that employing dynamic TDD with UAB can offer a uniform 30% gain in UL/DL mean rate over static TDD with SAB for all traffic scenarios captured by $\eta$. With no UE antenna gain, these gains are expected to be even higher as the access links become much weaker than backhaul. *In conclusion, one can harness the gains from dynamic TDD and UAB only if backhaul links are strong enough.* In the future, it would be desirable to develop analytical models that allow different antenna gains and path loss for backhaul links which would likely make dynamic TDD and UAB appear more favourable.

## 4.7 Summary

This is the first comprehensive study of UL-DL SINR distribution and mean rates in dynamic TDD enabled mmWave cellular networks. A key analytical takeaway is how to explicitly incorporate TDD frame structures for




(a) Downlink (b) Uplink

Figure 4.11: Dynamic TDD with UAB gives 30% gains over Static TDD with SAB in the noise-limited scenario at 73 GHz.

resource allocation studies in self-backhauled cellular networks using stochastic geometry. Computing approximate yet fairly accurate Laplace transform of new types of interference that arise while studying dynamic TDD and UAB is another takeaway with variety of applications. It can be useful to study co-existence of device-to-device/Internet-of-Things applications with cellular networks, wherein unscheduled UEs operate on the same band but for non-cellular purposes.

From a system insights viewpoint, the key takeaways lie in the comparison of different TDD schemes as a function of different access-backhaul splits, UL/DL traffic asymmetry and the density of BSs. Dynamic TDD and UAB are intriguing as they address some key fallacies with conventional static TDD and SAB implementations, as highlighted in this work, and it is worth noting that these are in fact a class of scheduling policies. The pros and cons of our heuristic implementations are exposed using the derived formulae under

various network settings, and the observations arouse interest in their further investigation with more sophisticated traffic models, implementation of self-backhauling with much *stronger* backhaul links than the access links and more realistic deployment and propagation assumptions. In the future, different variations of the class of scheduling policies considered in this work can be studied. For example, instead of employing dynamic TDD by considering only variation in UL and DL load per BS, it can be employed using the information on asymmetric in rates on UL and DL.

## 4.8 Appendices

### 4.8.1 Association probabilities

From Lemma 1 in [112], for $t \in \{m, s\}$ the CCDF of $\min_{X\in\Phi_t} L(X,0)$ is given by $V_t(\tau) = \mathbb{P}\left(\min_{X\in\Phi_t} L(X,0) > \tau\right) = \exp\left(-\Lambda_t(\tau)\right)$, where

$$\Lambda_t(\tau) = \pi\lambda_t\Bigg( \left(p_{\text{LOS}}\tau^{\frac{2}{\alpha_l}} + (1-p_{\text{LOS}})\tau^{\frac{2}{\alpha_n}}\right) \mathbb{1}(\tau < D_{\text{LOS}}^{\alpha_l}) + \tau^{\frac{2}{\alpha_n}}\mathbb{1}(\tau > D_{\text{LOS}}^{\alpha_n}) + \left(p_{\text{LOS}}D_{\text{LOS}}^2 + (1-p_{\text{LOS}})\tau^{\frac{2}{\alpha_n}}\right)\mathbb{1}(D_{\text{LOS}}^{\alpha_l} \le \tau \le D_{\text{LOS}}^{\alpha_n})\Bigg). \quad (4.14)$$

Here, $\Lambda_t(\tau)$ is the intensity of the propagation process $\{L(X,0) : X \in \Phi_t\}$. The PDF of $\min_{X\in\Phi_t} L(X,0)$ is given by $v_t(\tau) = \frac{\mathrm{d}\Lambda_t(\tau)}{\mathrm{d}\tau}\exp\left(-\Lambda_t(\tau)\right)$, where $\frac{\mathrm{d}\Lambda_t(\tau)}{\mathrm{d}\tau} =$

$$\frac{2\pi\lambda_t\tau^{\frac{2}{\alpha_n}-1}}{\alpha_n}\Bigg(\left(\frac{\alpha_n p_{\text{LOS}}\tau^{\frac{2}{\alpha_l}-\frac{2}{\alpha_n}}}{\alpha_l} + 1 - p_{\text{LOS}}\right)\mathbb{1}(\tau < D_{\text{LOS}}^{\alpha_l}) + (1-p_{\text{LOS}})\mathbb{1}(D_{\text{LOS}}^{\alpha_l} \le \tau \le D_{\text{LOS}}^{\alpha_n}) + \mathbb{1}(\tau > D_{\text{LOS}}^{\alpha_n})\Bigg). \quad (4.15)$$

Define, $\Lambda_t(\mathrm{d}\tau) = \frac{\mathrm{d}\Lambda_t(T)}{\mathrm{d}T}\big|_{T=\tau}\mathrm{d}\tau$ which will be useful in the Appendix 4.8.3. The probability that a typical user at origin associates with a MBS is given by

$$\begin{aligned}\mathcal{A}_m =& \mathbb{P}\left(\max_{X\in\Phi_m} \mathrm{P}_m L(X,0)^{-1} G_m B_m > \max_{Y\in\Phi_s} \mathrm{P}_s L(Y,0)^{-1} G_s B_s\right)\\ =& \int_0^\infty V_s\left(\frac{\mathrm{P}_s G_s B_s \tau}{\mathrm{P}_m G_m B_m}\right) v_m(\tau)\mathrm{d}\tau.\end{aligned}$$

If $\mathrm{P}_s G_s B_s = \mathrm{P}_m G_m B_m$, $\mathcal{A}_m = \lambda_m/\lambda_b$.

### 4.8.2 Proof of Lemma 8

The CDF of $\gamma_{a,D,X}\mathrm{F}_a$ is derived as follows. $\mathbb{P}\left(\gamma_{a,D,X}\mathrm{F}_a > r\middle|\mathrm{F}_a\right) = \mathbb{P}\left(\mathbb{1}(N_{d,X}>0)\frac{N_{d,X}\mathrm{F}_a}{N_{d,X}+N_{u,X}} > r\middle|\mathrm{F}_a\right) = p_1(r)$, which is computed using Assumption 2. Similarly, $\mathbb{P}\left(\gamma_{a,D,X}\mathrm{F}_a \geq r\middle|\mathrm{F}_a\right) = p_2(r)$ is derived. Now, let us denote $\tilde{\gamma}_{a,D,X} = \gamma_{a,D,X}\mathrm{F}_a - \lfloor\gamma_{a,D,X}\mathrm{F}_a\rfloor$. Thus,

$$\begin{aligned}\mathbb{P}\left(\mathrm{F}_{ad,D,X} = n\middle|\mathrm{F}_a\right) =& \mathbb{E}\left[\tilde{\gamma}_{a,D,X}\mathbb{1}(\lceil\gamma_{a,D,X}\mathrm{F}_a\rceil = n)\right.\\ &\left.+(1-\tilde{\gamma}_{a,D,X})\,\mathbb{1}(\lfloor\gamma_{a,D,X}\mathrm{F}_a\rfloor = n)\middle|\mathrm{F}_a\right]\\ =& \mathbb{E}\left[\tilde{\gamma}_{a,D,X}\mathbb{1}(n-1<\gamma_{a,D,X}\mathrm{F}_a\leq n)\right.\\ &\left.+(1-\tilde{\gamma}_{a,D,X})\,\mathbb{1}(n\leq\gamma_{a,D,X}\mathrm{F}_a<n+1)\middle|\mathrm{F}_a\right] \triangleq \mathbb{E}\left[\Xi\middle|\mathrm{F}_a\right].\end{aligned}$$

Since, $1\geq\Xi\geq 0$ the expectation can be computed as $\mathbb{E}\left[\Xi\middle|\mathrm{F}_a\right] = \int_0^1 \mathbb{P}\left(\Xi > r\middle|\mathrm{F}_a\right)\mathrm{d}r$. For $r=1$, the probability inside the integral is zero and for $r<1$,

$$\begin{aligned}&\mathbb{P}\left(\Xi > r\middle|\mathrm{F}_a\right) = \mathbb{P}\left(n+r-1<\gamma_{a,D,X}\mathrm{F}_a<n+1-r\middle|\mathrm{F}_a\right)\\ &= p_1(n+r-1) - p_2(n+1-r).\end{aligned}$$

### 4.8.3 Laplace functional of interference for computing access UL SINR

**Approximation 1:** Interference from MBS, SBS and UE is assumed independent of each other. Thus, $L_{i,w}^{ul,a,t,\mu}(\mathsf{s}, R) \approx$

$$\prod_{\nu\in\{m,s,u\}} \mathbb{E}\left[\exp\left(-\mathsf{s}I_{i,\nu,w}(X^*)\right)\middle| X^* \in \Phi_{t,\mu}, ||X^*|| = R, \mathcal{F}\right]$$
$$= L_m L_s L_u.$$

#### 4.8.3.1 Case 1: $i \leq \mathrm{F}_a$

**Interference from MBSs and SBSs.** This is non-zero only with dynamic TDD for access subframe. For $\nu \in \{m, s\}$ the Laplace transform can be simplified as follows. By superposition of PPPs, $\Phi_\nu = \Phi_{\nu,l} + \Phi_{\nu,n}$, wherein both the child processes are independent non-homogeneous PPPs with intensities $\lambda_\nu p_{\mathrm{LOS}} \mathbb{1}(x \leq D_{\mathrm{LOS}})$ and $\lambda_\nu(1 - p_{\mathrm{LOS}}\mathbb{1}(x \leq D_{\mathrm{LOS}}))$. Further, by strong Markov property of PPPs, replacing the shot noise of interference by that from independent copies of the PPPs, the following is derived.

$$L_\nu = \mathbb{E}\left[\exp\left(-\mathsf{s}\sum_{\mu_1\in\{l,n\}}\sum_{Y\in\Phi_{\nu,\mu_1}} \mathbb{1}(i \leq \mathrm{F}_{ad,w,Y}, N_{d,Y} > 0)\mathbb{1}\left(||Y||^{\alpha_{\mu_1}} > R^{\alpha_\mu}\frac{\mathrm{P}_\nu G_\nu B_\nu}{\mathrm{P}_t G_t B_t}\right) \times \right.\right.$$
$$\left.\left. \mathrm{C}_0 \mathrm{P}_\nu h_{i,X^*,Y} G_{i,X^*,Y} L(X^*, Y)^{-1}\right)\middle| X^* \in \Phi_{t,\mu}, ||X^*|| = R, \mathcal{F}\right]$$
$$= \mathbb{E}\left[\exp\left(-\mathsf{s}\sum_{\mu_1\in\{l,n\}}\sum_{\mu_2\in\{l,n\}}\sum_{Y\in\Phi_{\nu,\mu_1,\mu_2}} \mathbb{1}(i \leq \mathrm{F}_{ad,w,Y})\mathbb{1}\left(||Y||^{\alpha_{\mu_1}} > R^{\alpha_\mu}\frac{\mathrm{P}_\nu G_\nu B_\nu}{\mathrm{P}_t G_t B_t}\right) \times \right.\right.$$
$$\left.\left. \mathbb{1}(N_{d,Y} > 0)\mathrm{C}_0 \mathrm{P}_\nu h_{i,X^*,Y} G_{i,X^*,Y} ||X^* - Y||^{-\alpha_{\mu_2}}\right)\middle| X^* \in \Phi_{t,\mu}, ||X^*|| = R, \mathcal{F}\right].$$

where $\Phi_{\nu,\mu_1,\mu_2}$ are BSs of tier $\nu$ which have type $\mu_1 \in \{l,n\}$ links to the origin and type $\mu_2 \in \{l,n\}$ links to $X^*$. Given, $||X^*|| = R$, $\Phi_{\nu,\mu_1,\mu_2}$ is a PPP with density $\hat{\lambda}_{\nu,\mu_1,\mu_2}(r,\theta) = \lambda_\nu p_{\mu_1}(r) p_{\mu_2}\left(\sqrt{r^2+R^2-2rR\cos(\theta)}\right)$. Further simplifying, the above expression is equal to

$$\prod_{\mu_1,\mu_2} \exp\left(-\int_{\left(\frac{R^{\alpha_\mu} \mathrm{P}_\nu G_\nu B_\nu}{\mathrm{P}_t G_t B_t}\right)^{1/\alpha_{\mu_1}}}^{\infty} \int_0^{2\pi} \mathbb{E}\left[\frac{p_{i,w,\nu}\hat{\lambda}_{\nu,\mu_1,\mu_2}(r,\theta) r}{1+\frac{(r^2+R^2-2rR\cos(\theta))^{\alpha_{\mu_2}/2}}{\mathsf{s}\mathrm{C}_0 \mathrm{P}_\nu \Psi_{t,\nu}}}\right] \mathrm{d}r\mathrm{d}\theta\right),$$

where

$$p_{i,w,\nu} = \mathbb{P}\left(N_d > 0, i \leq \mathrm{F}_{ad,w} \middle| \mathcal{F}\right) = \sum_{n=i}^{\mathrm{F}_a} \mathbb{P}\left(\mathrm{F}_{ad,D} = n \middle| \mathcal{F}\right), \tag{4.16}$$

which can be computed using Lemma 8.

Note that the lower limit of integral on $r$ is exactly the value of $\mathsf{s}$ from (4.11). Thus, rewriting the equation with change of variables $\rho = r\left(\frac{R^{\alpha_\mu} \mathrm{P}_\nu G_\nu B_\nu}{\mathrm{P}_t G_t B_t}\right)^{-1/\alpha_{\mu_1}}$ is easier to implement on MATLAB. An even easier implementation, which is in fact a lower bound to the Laplace functional, can be obtained by neglecting the $\mathbb{1}(||Y||^{\alpha_{\mu_1}} > (.))$ term in the above derivation, which gives lower bound in Lemma 9.

**Interference from UEs.** $\mathbb{E}[\exp(-\mathsf{s}I_{i,u,w}) \,|X^* \in \Phi_{t,\mu}, ||X^*|| = R, \mathcal{F}]$ can be computed using a non-homogeneous PPP approximation inspired from [107].

**Approx. 2.** *Laplace transform of interference from scheduled device process ($\Phi_1$) connected to a PPP BS process ($\Phi_2$) to a receiver under consideration is approximated by that generated from an independent PPP device process $\Phi_3$ with same intensity as $\Phi_2$. Further thinning is done $\Phi_3$ to approximate the*

*pair correlation function by taking into consideration the association of points in $\Phi_3$ to those in $\Phi_2$ [107].*

Thus, conditioned on the event that the tagged BS $X^*$ is of tier $t$, the propagation process of interfering UEs is approximately equal in distribution to an independent non-homogeneous PPP on $\mathbb{R}^+$ with intensity

$$\Lambda(t, \mathrm{d}r) = \sum_{k \in \{m,s\}} \left(1 - \exp\left(-\Lambda_k\left(r \frac{\mathrm{P}_k B_k G_k}{\mathrm{P}_t B_t G_t}\right)\right)\right) p_{i,w,k} \Lambda_k(\mathrm{d}r), \tag{4.17}$$

with $\Lambda_k(\mathrm{d}r) = \frac{\mathrm{d}\Lambda_k(x)}{\mathrm{d}x}\Big|_r \mathrm{d}r$, and $p_{i,w,k} = \mathbb{P}\left(N_u > 0, \mathrm{F}_{ad,w} < i \leq \mathrm{F}_a \middle| \mathcal{F}\right)$. Note that $p_{i,w,k}$ is captures the active probability of interferer in the $i^{\text{th}}$ slot and the non-idle probability of parent BS process. The $1 - \exp(.)$ term ensures that the biased received power from at least one of the points in $\Phi_k$ is better than that from the BS at $X^*$ [107]. Thus,

$$L_u \approx \exp\left(-\int_0^\infty \mathbb{E}\left[\frac{1}{1 + \frac{r}{\mathsf{s}\mathrm{C}_0 \mathrm{P}_u \Psi_{t,u}}}\right] \Lambda(t, \mathrm{d}r)\right). \tag{4.18}$$

Here, $p_{i,S,k} = \mathbb{P}\left(N_{u,X} > 0, \mathrm{F}_{ad,S,X} < i \leq \mathrm{F}_a \middle| \mathcal{F}\right) =$

$$\left(1 - \left(1 + \frac{\lambda_u \mathcal{A}_k (1-\eta)}{3.5 \lambda_k}\right)^{-3.5}\right) \mathbb{1}\left(\mathrm{F}_{ad} < i \leq \mathrm{F}_a\right).$$

Since, an UL UE is only scheduled in access subframe for $\mathrm{F}_{ad} < i \leq \mathrm{F}_a$ with static TDD, the indicator in previous expression will always be 1 for feasible UL access SINR distributions. Similarly, $p_{i,D,k}$

$$= \mathbb{P}\left(\mathrm{F}_{ad,w} < i \leq \mathrm{F}_a \middle| \mathcal{F}\right) - \mathbb{P}(N_u = 0, \mathrm{F}_{ad,w} < i \leq \mathrm{F}_a | \mathcal{F})$$
$$= \mathbb{P}\left(\mathrm{F}_{ad,w} < i \leq \mathrm{F}_a \middle| \mathcal{F}\right) - \mathbb{P}(N_u = 0 \middle| \mathcal{F})$$

The first term can be found by substituting $t = k$ in Lemma 8 and the second term is $\left(1 + \frac{\lambda_u \mathcal{A}_k (1-\eta)}{3.5\lambda_k}\right)^{-3.5}$.

#### 4.8.3.2 Case 2: $i > \mathrm{F}_a$ and $w = \mathbf{UAB}$

Note that if we are computing Laplace functional of interference at an UL receiver of an access link for $i > \mathrm{F}_a$, by definition we are operating in $w = \mathrm{UAB}$ mode with $X^* \in \Phi_s$. In this case there is no interference from MBSs.

**Interference from SBSs.** The interference from SBSs can be computed similar to the previous case on interfering UEs with $i < \mathrm{F}_a$. However, we need to incorporate the fact that the MBS serving $X^*$ has an interfering SBS scheduled with probability 1 but other MBSs may not have a scheduled SBS with probability $p_{i,w,s} = \left(1 + \frac{\lambda_{s,u}}{3.5\lambda_m}\right)^{-3.5}$ with $\lambda_{s,u} = \lambda_s \left(1 - \left(1 + \frac{\mathcal{A}_s \lambda_u (1-\eta)}{3.5\lambda_s}\right)^{-3.5}\right)$. Thus, the following version of approx. 2 is employed. The point closest to $X^*$ in the new interfering PPP is active with probability 1 and rest of the points are active with probability $p_{i,w,s}$. This gives the corresponding expression in Lemma 10.

**Interference from UEs.** By approximation 2, the interfering PPP process has intensity equal to $\lambda_s$. A further thinning by $\frac{\hat{\lambda}}{\lambda_s}$ is done, where $\hat{\lambda} =$

$$\mathbb{1}\left(\mathrm{F}_a + \mathrm{F}_{bd} < i \leq \mathrm{F}\right) \left(1 - \left(1 + \frac{\lambda_u (1-\eta) \mathcal{A}_s}{3.5\lambda_s}\right)^{-3.5}\right)$$

$$\times\, p_{ul}\left(\lambda_s - \left(1 - \left(1 + \frac{\lambda_s}{3.5\lambda_m}\right)^{-3.5}\right)\lambda_m\right)^{+},$$

where $a^+ = a$ if $a > 0$ and zero otherwise. This captures that there will be at most 1 scheduled UE from every SBS with poaching probability $p_{ul}$ except those SBSs which are scheduled by their serving MBS. Thus, the Laplace functional is same as (4.18) but with $\Lambda(t, \mathrm{d}r)$ replaced by $\frac{\hat{\lambda}}{\lambda_s}\left(1 - \exp\left(-\Lambda_s(r)\right)\right)\Lambda_s(\mathrm{d}r)$, where 1-exp(.) accounts for the probability that the interfering UEs don't associate with the SBS at $X^*$.

### 4.8.4 Uplink mean rate

In the following derivation of UL mean rate, $w_a \in \{S, D\}$ and $w_b \in \{\mathrm{SAB}, \mathrm{UAB}\}$.

$$\begin{aligned}
\mathrm{R}_{ul,m,w_a} &= \frac{\mathbb{E}\left[\mathrm{D}_{ul,m,w_a}\middle|\mathcal{E}_m\right]}{\mathrm{TF}} \\
&= \frac{\mathrm{W}}{\mathrm{F}}\mathbb{E}\left[\sum_{i=1+\mathrm{F}_{ad,w_a,X^*}}^{\mathrm{F}_a} \mathbb{1}\left(\text{UE scheduled in } i^{\text{th}} \text{ slot}\right)\log_2\left(1+\mathsf{SINR}^{ul}_{i,a,w_a}\right)\middle|\mathcal{E}_m\right] \\
&= \frac{\mathrm{W}}{\mathrm{F}}\mathbb{E}\left[\sum_{i=1+\mathrm{F}_{ad,w_a,X^*}}^{\mathrm{F}_a} \frac{\mathbb{E}\left[\log_2\left(1+\mathsf{SINR}^{ul}_{i,a,w_a}\right)\middle|\mathrm{F}_a, N_{u,X^*}, N_{d,X^*}, \mathcal{E}_m\right]}{N_{u,X^*}}\middle|\mathcal{E}_m\right] \\
&= \frac{\mathrm{W}}{\mathrm{F}}\mathbb{E}\left[\sum_{i=1+\mathrm{F}_{ad,w_a,X^*}}^{\mathrm{F}_a} \frac{1}{N_{u,X^*}}\int_0^\infty \frac{\mathsf{S}^{ul,m}_{i,a,w_a}(\tau)}{1+\tau}\mathrm{d}\tau\middle|\mathcal{E}_m\right],
\end{aligned}$$

where distribution of $\mathrm{F}_{ad,D,X^*}$ given $\gamma_{a,D,X} = \frac{n_2}{n_1+n_2+1}$ is given by (4.3). Similarly, given the constant $\gamma_a$ the distribution of $\mathrm{F}_{ad,S,X^*}$ can also be found from (4.3). To compute $\mathrm{R}_{ul,s,w_a,w_b}$, let us look at each of the expectations inside the

minimum one by one.

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{D}_{ul,s,a,w_a,w_b}\middle|\mathcal{E}_s\right] \\
&= \mathrm{WT}\mathbb{E}\left[\sum_{i=1+\mathrm{F}_{ad,w_a,X^*}}^{\mathrm{F}_a} \frac{1}{N_{u,X^*}} \int_0^\infty \frac{\mathsf{S}_{i,a,w_a}^{ul,s}(\tau)}{1+\tau}\mathrm{d}\tau \middle| \mathcal{E}_s\right] \\
&+ \mathbb{1}(w_b = \mathrm{UAB})\mathrm{WT}\mathbb{E}\left[\sum_{i=1+\mathrm{F}_a+\mathrm{F}_{bd}}^{\mathrm{F}} \left(1-\frac{1}{N_{s,X^{**}}}\right)\frac{p_{ul}}{N_{u,X^*}} \int_0^\infty \frac{\mathsf{S}_{i,a,w_b}^{ul,s}(\tau)}{1+\tau}\mathrm{d}\tau \middle| \mathcal{E}_s\right],
\end{aligned}$$

where $N_{s,X^{**}}$ is the number of SBSs associated with $X^{**}$ with at least one UL UE. Similarly,

$$\mathbb{E}\left[\mathrm{D}_{ul,s,b,w_b}\middle|\mathcal{E}_s\right] = \mathrm{WT}\mathbb{E}\left[1/N_{s,X^{**}}\right]\sum_{n=1}^{\infty}\frac{\kappa_{s,ul}^*(n)}{n}\mathbb{E}_{\mathcal{F}}\int_0^\infty \frac{\sum_{i=1+\mathrm{F}_a+\mathrm{F}_{bd}}^{\mathrm{F}}\mathsf{S}_{i,b,w_b}^{ul,s}(\tau)}{1+\tau}\mathrm{d}\tau.$$

### 4.8.5 Laplace functional of interference for access DL SINR

The main difference with UL case is that now the receiver is at origin instead of at $X^*$. Thus, different exclusion regions need to be considered while computing the shot noise. By approximation 1,

$$\begin{aligned}
&L_{i,w}^{dl,a,t,\mu}(\mathsf{s},R) \\
&\approx \prod_{\nu\in\{m,s,u\}} \mathbb{E}\left[\exp\left(-\mathsf{s}I_{i,\nu,w}(0)\right)\middle|X^*\in\Phi_{t,\mu},||X^*||=R,\mathcal{F}\right].
\end{aligned}$$

**Case 1:** $i \le \mathrm{F}_a$**.** For $\nu \in \{m,s\}$,

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(-\mathsf{s}I_{i,\nu,w}(0)\right)\middle|X^*\in\Phi_{t,\mu},||X^*||=R,\mathcal{F}\right] \\
&= \exp\left(-\int_{R^{\alpha_\mu}}^\infty \mathbb{E}\left[\frac{1}{1+\frac{r}{\mathsf{s}\mathrm{C}_0\mathrm{P}_\nu\Psi_{t,u}}}\right]p_{i,w,d}\Lambda_\nu(\mathrm{d}r)\right),
\end{aligned}$$

where $p_{i,w,\nu}$ is given (4.16) and $\Lambda_\nu(\mathrm{d}r)$ was defined in Appendix 4.8.1. Note that this is exact expression.

For $\nu = u$, there will non-zero interference only with dynamic TDD. By approximation 2, we compute the Laplace functional of interference from UEs is generated from two independent PPPs – for SBS/MBS connection – as follows.

$$\begin{aligned}&\mathbb{E}\left[\exp\left(-\mathsf{s}I_{i,u,w}(0)\right)\middle|X^*\in\Phi_{t,\mu},||X^*||=R,\mathcal{F}\right]\\&\approx\exp\left(-\int_0^\infty\mathbb{E}\left[\frac{1}{1+\frac{r}{\mathsf{s}\mathrm{C}_0\mathrm{P}_u\Psi_{u,u}}}\right]\sum_{k\in\{m,s\}}p_{i,w,k}\Lambda_k(\mathrm{d}r)\right),\end{aligned}$$

where $p_{i,w,k}$ can be found just after (4.18).

**Case 2:** $i > \mathrm{F}_a$**.** In backhaul subframe, a DL UE is scheduled for access only if $\mathrm{F}_a < i \leq \mathrm{F}_a + \mathrm{F}_{bd}$, $w$ =UAB and the UE connects to a SBS. Thus, there is interference only from MBSs and SBSs.

$$\begin{aligned}&\mathbb{E}\left[\exp\left(-\mathsf{s}I_{i,m,w}(0)\right)\middle|X^*\in\Phi_{s,\mu},||X^*||=R,\mathcal{F}\right]\\&=\exp\left(-\int_{R^{\alpha_\mu}}^\infty\mathbb{E}\left[\frac{1}{1+\frac{r}{\mathsf{s}\mathrm{C}_0\mathrm{P}_m\Psi_{m,u}}}\right]p_{i,w,m}\Lambda_m(\mathrm{d}r)\right),\end{aligned}$$

where

$$p_{i,w,m}=\mathbb{1}(\mathrm{F}_a<i\leq\mathrm{F}_a+\mathrm{F}_{bd})\left(1-\left(1+\frac{\lambda_{s,d}}{3.5\lambda_m}\right)^{-3.5}\right)$$

with $\lambda_{s,d}=\lambda_s\left(1-\left(1+\frac{\lambda_u\eta_\mathcal{A}s}{3.5\lambda_s}\right)^{-3.5}\right)$.

To compute $\mathbb{E}\left[e^{-\mathsf{s}I_{i,s,w}(0)}\middle|X^*\in\Phi_{s,\mu},||X^*||=R,\mathcal{F}\right]$, we make the following approximation similar to the corresponding UL case for poaching. The

SBS interferers form an independent homogeneous $\Phi_{\mathrm{BS}}$ with density given by

$$\hat{\lambda}_d = \left(\lambda_s - \left(1 - \left(1 + \frac{\lambda_s}{3.5\lambda_m}\right)^{-3.5}\right)\lambda_m\right)^+ \times \\ p_{dl}\mathbb{1}\left(\mathrm{F}_a < i \leq \mathrm{F}_a + \mathrm{F}_{bd}\right)\left(1 - \left(1 + \frac{\lambda_u \eta \mathcal{A}_s}{3.5\lambda_s}\right)^{-3.5}\right).$$

Thus, we get

$$\mathbb{E}\left[\exp\left(-\mathsf{s}I_{i,s,w}(0)\right)\middle| X^* \in \Phi_{s,\mu}, ||X^*|| = R, \mathcal{F}\right] \\ \approx \exp\left(-\int_0^\infty \mathbb{E}\left[\frac{1}{1 + \frac{r}{\mathsf{s}\mathrm{C}_0\mathrm{P}_s\Psi_{s,u}}}\right]\frac{\hat{\lambda}_d}{\lambda_s}\Lambda_s(\mathrm{d}r)\right).$$

# Chapter 5

# How Many Hops Can Self-backhauled Millimeter Wave Cellular Networks Support?[10]

In order to deploy affordable millimeter wave (mmWave) cellular networks, it is highly desirable to deploy self-backhauled networks. In the previous chapter, we studied mean per user rates in mmWave self-backhauled networks with a single backhaul hop. This chapter addresses the following key question for designing financially viable mmWave cellular networks. What is the maximum extended coverage area that a single fiber site can support using multi-hop relaying, while still achieving a minimum target per user data rate? We formulate an optimization problem to maximize the minimum end-to-end per user data rate, and exploit unique features of millimeter wave deployments to yield a tractable solution.

Although mesh network architectures have been considered both the-

[10]This chapter reproduces the content of the following publication. M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "How Many Hops Can Self-backhauled Millimeter Wave Cellular Networks Support?", submitted to *IEEE Trans. Wireless Commun.* in May 2018. The research performed in this chapter including setting up the system model, the formulation of the analytical problem, solving it and generating all numerical results are primarily my contribution. My co-authors, J. G. Andrews and A. Ghosh, guided me in identifying the research problem, and they also gave me regular feedback while I was working on the problem, and while I was writing the paper. I would like to thank G. de Veciana for his suggestions on formulating the max-min rate optimization problem as in this work.

oretically and in practice many times in the past (as discussed next), with limited success, a few novel features of urban mmWave cellular systems lends to significant simplification. In particular, the highly directional transmissions, strong blocking from buildings, and limited diffraction around corners [21, 32, 113, 114] – combined with an urban topography – allow us to plausibly model the network as a noise-limited $k-$ring deployment model, as shown in Fig. 5.1, with BSs deployed on a 2-D square grid. The number of relays grows as $k^2$ with a fixed inter-site distance (ISD). We consider a single fiber site, ignoring edge effects, which are anyway negligible due to the noise-limitedness. This model will allow us to succinctly quantify the maximum achievable rates by all users, called max-min rates, in closed form.

We focus on max-min rates for two reasons. The first is that it allows us to determine the maximum value of $k$, that is how far the mesh network can extend from the fiber site, while ensuring a certain end-to-end (e2e) data rate. The second is that it results in a tractable optimization problem, as opposed to focusing on say, the 5th percentile user. We provide several validations of the proposed model and results. Given these tractable results, we consider three additional design choices, namely (i) integrated access-backhaul (IAB) or orthogonal access-backhaul (OAB) resource allocation, (ii) full or half duplex relays, and (iii) does dual connectivity improve per user rates in a self-backhauled network? IAB allows access and backhaul links to share time-frequency resources, whereas OAB reserves different set of resources for access and backhaul links.

## 5.1 Background, Motivation, and Related Work

The study of multi-hop wireless networks has a rich history spanning theoretically optimal resource allocation schemes [115–118], scaling laws [119–121] and analysis of achievable e2e metrics [122–126]. There has also been industry-driven standardization activities for multi-hop wireless local area networks (WLANs) [127] and for fourth generation (4G) cellular networks with a single wireless backhaul hop [128, 129]. Practical implementation of multi-hop networks, however, has not been very successful. Reasons include the coupled interference and scheduling between hops [124], large overheads for maintaining multi-hop routes, a lack of Shannon-like theoretical limits and their corresponding design guidance [130] and the fundamentally poor e2e-rate scaling caused by each packet having to be transmitted multiple times [119].

A key differentiating factor for mmWave cellular networks is that they often tend to be noise-limited, especially for large bandwidths and small antenna beam widths [46, 131]. This noise-limitedness greatly simplifies the routing and multi-hop scheduling problems and helps to close the gap between theoretically optimal solutions and practical implementations. For example, recently [45] proposes a polynomial time algorithm for joint routing and scheduling, extending the work in [115], unlike traditional NP-hard solutions [116, 118]. However, [45] considers a generic deployment topology and exploiting specific deployment patterns may result in even simpler solutions to optimal routing or scheduling. For example, in this paper we prove (for specific load scenarios) the optimality of a static routing scheme for urban canyon de-

ployments, called *nearest neighbour highway routing*. Additionally, we observe it to be optimum more generally using empirical studies. We also observe that because a few links are critical bottlenecks for e2e rate, it is usually sufficient for an optimal scheduler to activate only a few links at a time. Another reason in favor multi-hop mmWave cellular networks is that many high data rate 5G applications have limited or no mobility, such as fixed wireless-to-home, industrial automation, or pedestrian/static mobile broadband users [9]. For future mobile applications, multi-connectivity, where users connect to multiple BSs at a time potentially operating over different frequency bands, may be exploited to offer smooth handovers in mmWave networks [132].

This new-found interest in multi-hopping for mmWave is reflected in recent work such as [40–46, 133]. A cross layer optimization framework was proposed in [40]. In [133], a fixed demand per flow traffic model was assumed to solve the problem of minimizing the time to empty the demands of all flows. In [41], a joint cost minimization along with resource allocation optimization problem was formulated. In [46], per user rate analysis in mmWave self-backhauled networks was done using stochastic geometry considering a single backhaul hop. This framework does not trivially extend to analysis of multi-hop backhauling. Note that in most of the prior works which attempted to optimize resources in multi-hop mmWave networks the optimal solutions are NP-hard or require implementing a linear program that involve matrices whose size grows very fast with size of the network (usually involves finding maximal matchings of a graph to list all possible valid schedules under half

duplex constraint) [42, 45], although approximate simpler solutions have been attempted [44, 45, 133]. The closest recent work to this work is [43], which considers a general graph for deployment and includes out of cell interference leading to a linear programming solution for joint routing and scheduling. This is where our work differs, as we use noise-limitedness to exactly solve the optimization problem under consideration to give closed form results for maximizing minimum rate in the network considering different design choices and also provide structural results on optimal routing. We later show the utility of these results by comparing with empirical studies incorporating interference.

## 5.2 Contributions

**Closed form results for max-min optimal rates.** We propose to study a grid deployment of BSs with a single fiber site and $k^2$ relays around it, which we term a $k-$ring deployment. Arbitrary but static UE deployment is assumed with full buffer traffic. UEs can be uplink (UL) or downlink (DL). We compute closed form expressions of maximum e2e rate achievable by all UEs when the BSs are half or full duplex, and when IAB or OAB resource allocation strategy is used. All rates are assumed to be deterministic by default, although our results for OAB hold when access rates are random. We first compute the max-min rates with simplifying assumptions on load across different BSs, and assume equal access rates for all users to come up with a simple formula. Optimality of nearest neighbor highway routing, defined in Section 5.4, is shown in this scenario but is later observed to hold in greater generality. The

analytical result is then extended considering a more general setup of unequal access rates to different users, arbitrary load per BS. The proof technique used in this work is to first find an upper bound to max-min rates and then it is shown that the upper bound is achievable.

**Applications of the analysis.** We answer the following question under several realistic network parameters for BSs spaced at 200 meters and operating at 28 GHz carrier frequency with 800 MHz bandwidth. How many hops can self-backhauled millimeter wave cellular networks support for meeting a minimum 100 Mbps target rate per user? If 1024 QAM is the maximum size of constellation supported, up to 4 ring deployments can offer 100 Mbps per UE considering a load of 2 UEs per BS. If even higher order modulations are used, up to 6 ring deployments can be supported with 2 UEs per BS for practical values of antenna gains and D = 100m. The max-min rates derived are also used to compare IAB versus OAB, and half versus full duplex base stations, which also lead to interesting insights detailed in Section 5.6. For instance, it is possible to closely follow the max-min rates with IAB using an OAB scheme which can be simpler to implement in practice.

**Positive side-effects of network bottlenecks.** We observe that our noise-limited analysis is accurate not just because of large bandwidth or narrow beam-widths, but also for the following reason. The proposed deployment has very few bottleneck links in several load scenarios considering reasonably large antenna gains (greater than 16 UE antennas and 64 BS antennas [16]) such that most NLOS UE access links are not bottlenecks. Thus, the optimal scheduler

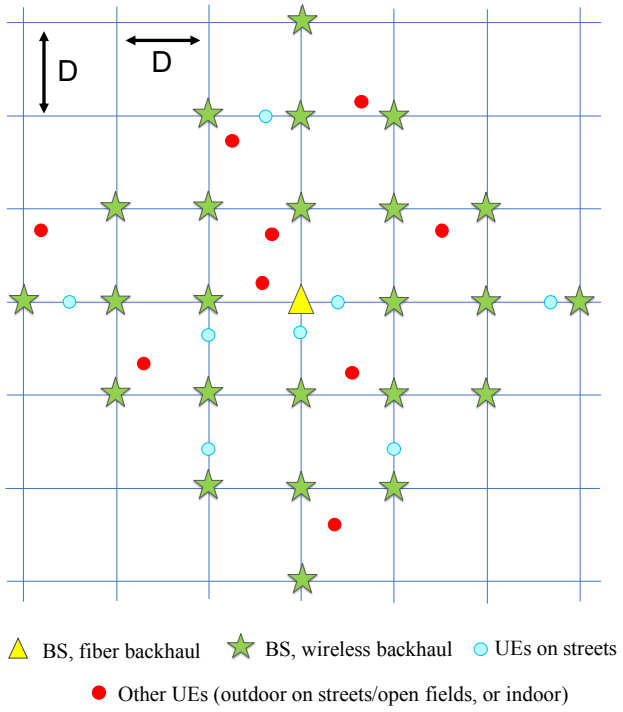



Figure 5.1: $k-$ring model, $k = 3$.

can meet the theoretically optimal max-min rates by just activating few links at a time, leading to noise-limited system performance. Another positive side-effect of network bottlenecks is that the analysis can be used as a benchmarking tool for complex simulators which emulate proportional fairness (PF) in multi-hop networks. We show an illustrative example which also highlights that $k-$ring deployments can be noise-limitedness even if the schedulers are not interference aware, owing to blockage effects and directionality of mmWave networks.

## 5.3 System Model

**$k-$ring deployment.** We propose to study a $k-$ring deployment model for urban canyon scenarios, as shown in Fig. 5.1. Lines represent streets on which BSs are deployed, with BSs denoted by either a triangle (MBS) or star (relays). The inter-line spacing is D meters. The MBS, which is the fiber backhauled BS or master BS, is located at $(0, 0)$ and the relays are located at $(i\text{D}, j\text{D})$ for $i, j \in \{0, \pm 1, \pm 2, \ldots, \pm k\}$ such that the Manhattan distance from any relay to the MBS is $\leq k\text{D}$. For simplicity, we denote $(i\text{D}, j\text{D})$ by $(i, j)$. BSs separated by a distance D are LOS. Fig. 5.2 shows a heuristic $k-$ring deployment ($k = 4$, D = 100m) in Chicago's downtown, which indicates it is a reasonable model. All possible links (directed line joining any two nodes, which can be BSs or UEs) in the network are wireless.

Performance is evaluated for a static realization of UE locations, motivated from fixed wireless to home or other low mobility applications. The analysis in Section 5.4 works for arbitrary UE locations, and specific assumptions on access rates to different UEs will be made later while discussing the results. Let U be the total number of UEs. Each UE associates with one BS according to a any association criterion, which does not change with time. For example, nearest neighbour or minimum path loss association. UEs can only connect with BSs along the same street since path loss on links across orthogonal streets can be very high [114, 134]. Number of users connected to a BS at $(i, j)$ is denoted as $w_{i,j}$. A downlink (DL) network is assumed, which implies that the fiber site transmits to all the UEs via relays. BSs and UEs

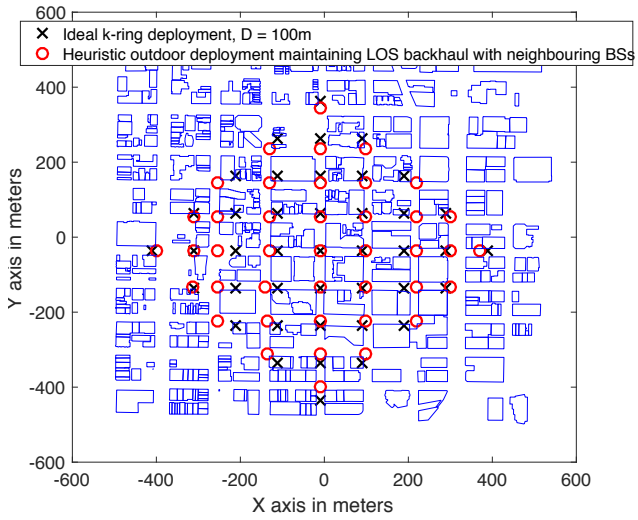



Figure 5.2: $k-$ring model overlaid in an urban area with D = 100m.

are assumed to employ single stream beamforming. All devices in the network are assumed to be half-duplex, unless explicitly stated to be full duplex.

**Routing and traffic model.** Time is assumed to be continuous with no explicit slot structure. Total time is unit. An ordered list of all nodes visited by a UE's data starting from the MBS is called the *route* of that UE. The route includes the UE itself. A *hop* on the route of a UE is a link between adjacent nodes in the route of the UE. It is assumed that backhaul communication (that is BS to BS hops) can happen on links only along the same street [114, 134]. Furthermore, it is assumed there is a *unique* route from the fiber site to every UE. However, different UEs associated with the same BS can have different routes. For instance, if there are two UEs connected to a BS at $(-1,-1)$, then one of the UE can have a route $(0,0) \rightarrow (-1,0) \rightarrow (-1,-1) \rightarrow \text{UE}_1$ and the other UE can have a route $(0,0) \rightarrow (0,-1) \rightarrow (-1,-1) \rightarrow \text{UE}_2$.

Full buffer traffic model is assumed. This implies that given a route of a UE, every BS along the route always has the UE's data to transmit. *Routing strategy* is defined as the collection of routes of all UEs. Given a routing strategy, $f(i,j)$ denotes the effective number of UEs served by $(i,j)$. That is, $f(i,j) = \sum_{u=1}^{\mathrm{U}} \mathbf{1}\{(i,j) \in \text{route of user } u\}$. Note that $f(0,0) = \mathrm{U}$.

**Instantaneous rate and noise-limitedness.** Every link (access and backhaul) in the network is associated with a fixed number called instantaneous rate. If a link with instantaneous rate $R$ is activated for time $\tau$, then $\tau R$ is the data transmitted on that link. Let $R_i$ denote the deterministic instantaneous rate on a backhaul link of length $i$D for $i = 1, \ldots, k$. It is assumed that $R_i$ is decreasing with $i$. Assumptions on instantaneous access rates will be made in the next section. Note that backhaul links along a street will generally be LOS. Since LOS mmWave links have negligible small scale fading [32], an assumption of deterministic instantaneous rates is justifiable. The analytical results with OAB can be extended for random instantaneous rates for access links, which can incorporate the impact of dynamic blockages, and we will discuss more about this later. Another implicit but important assumption was made above. That is, the instantaneous rates are independent of the *transmission schedules*, that is the set of links activated simultaneously. This is essentially noise-limitedness assumption, which we will extensively validate in Section 5.7.

**Scheduling assumptions.** Let L be the total number of links, then feasible schedules are defined by a collection of $\mathrm{L} \times \mathrm{U}$ matrices, called *scheduling*

*matrices*, which are described next. Each entry, $\tau_{l,u}$, in a scheduling matrix indicates fraction of time link $l$ was used to serve data for user $u$. Here, $0 \leq \tau_{l,u} \leq 1$. Furthermore, since the total time over which optimization is done is 1 unit, the fraction of time every BS is active (that is either transmitting or receiving) is 1. That is, $\sum_{l \in \mathcal{L}_{i,j}} \sum_{u=1}^{\mathrm{U}} \tau_{l,u} \leq 1$, where $\mathcal{L}_{i,j}$ denotes the set of links connected to $(i, j)$ given a routing strategy.

## 5.4 Max-min end to end rate in $k-$ring deployment

We want to study what is the maximum value of $k$ that can support a target e2e rate achieved by all UEs. We instead fix a $k$ and find the maximum e2e rate achieved by all UEs. Let us define this formally. *Long term rate* of a user $u$ on link $l$ is defined as $\tau R$, where $R$ is the instantaneous rate on link $l$ and $\tau$ ($< 1$) is the fraction of time that user $u$ was scheduled on link $l$. Given a routing strategy and a corresponding scheduling matrix $\mathcal{S}$, the *e2e rate of UE* $u$, denoted as $R_u^{\mathcal{S}}$, is the minimum of its long term rate over all hops from the fiber site to the UE.

**Definition 7.** Given a routing strategy, the max-min rate is defined as $\gamma^* = \max_{\mathcal{S}} \theta_{\mathcal{S}}$, where $\theta_{\mathcal{S}} = \min_{u=1,\ldots,\mathrm{U}} R_u^{\mathcal{S}}$ with $R_u^{\mathcal{S}}$ being the e2e rate of user $u$ for scheduling matrix $\mathcal{S}$. Maximizing $\gamma^*$ considering all routing strategies that are feasible as per the system model in Section 5.3, we obtain the globally optimal max-min rate denoted as $R_{\mathrm{e2e}}^*$.

Given a routing strategy, the max-min rate optimization problem can

be restated as follows.

$$\begin{aligned}
\underset{\mathcal{S}}{\text{maximize}} \quad & \theta_{\mathcal{S}} \\
\text{subject to} \quad & \theta_{\mathcal{S}} \leq r_l \tau_{l,u}, \forall \text{ hop } l \text{ on the route of user } u, \forall u = 1, \ldots, \mathrm{U}, \\
& \sum_{l \in \mathcal{L}_{i,j}} \sum_{u=1}^{\mathrm{U}} \tau_{l,u} \leq 1, \forall\, i, j \in \{0, \pm 1, \ldots, \pm k\} \text{ s.t. } |i| + |j| \leq k, \\
& 0 \leq \tau_{l,u} \leq 1, \forall\, l \in \{1, 2, \ldots, \mathrm{L}\}, u \in \{1, \ldots, \mathrm{U}\} \\
& \tau_{l,u} = 0, \text{ if link } l \text{ is not a hop on route of UE } u, \forall u \in \{1, \ldots, \mathrm{U}\},
\end{aligned}$$

where $r_l$ is the instantaneous rate on link $l$ and $\mathcal{S}$ denotes $\mathrm{L} \times \mathrm{U}$ matrix with elements $\tau_{l,u}$ for $l = 1, \ldots, \mathrm{L}$ and $u = 1, \ldots, \mathrm{U}$. Note that $R^*_{\mathrm{e2e}}$ is the maximum of the solution to the above problem over all feasible routing strategies (constraints have been described in Section 5.3).

We first analyze max-min e2e rates for IAB and then turn to analysis of OAB schemes. IAB allows access and backhaul to share time-frequency resources, whereas there is a fixed split between access and backhaul across all BSs with OAB.

### 5.4.1 Integrated access backhaul

We first derive max-min rates for a special case of UE load and access rates. Then we extend the result to a more general setup.

**Highway routing.** We consider a class of routing strategies called *highway routing*. This is defined as follows. Streets along the X and Y axes are called as *highways*. All UEs associated with a BS at $(i, j)$ have same route from the fiber site to the associated BS. Under a highway routing strategy,

the fiber site first transmits data to either $(i, 0)$ or $(0, j)$, whichever is furthest in terms of Manhattan distance, potentially over multiple hops. From $(i, 0)$ or $(0, j)$ the data is then transmitted to the $(i, j)$ along the shortest path in terms of Manhattan distance, potentially over multiple hops. The Manhattan distance of $(i, j)$ from the MBS decreases with every DL hop. If $|i| = |j|$, then the traffic is directed to either $(i, 0)$ or $(0, j)$. However, if $(0, 0) \to (i, 0) \to (i, i)$ then $(0, 0) \to (-i, 0) \to (-i, -i)$. If there were UL paths, then those would be exactly same as DL paths but in reverse order. Theorem 9 proves optimality of nearest neighbour highway routing (NNHR) in specific load scenarios and when access rates to all users is the same. We then discuss why NNHR is a good choice in more general load and access rate settings. In Section 5.7, we empirically observe that NNHR gives optimal performance in the general UE load scenarios under consideration.

For the first result, all access links are assumed to have a common instantaneous rate $R_a$. This can be thought as an outcome of power control or just a simplifying assumption. DL UEs are assumed for simplicity of exposition, although not necessary for the result to hold.

**Theorem 9.** *Let $w_{0,0} \geq w_{i,j}$ and $w_{i,j} = w_{-i,-j}$ $\forall i, j \in \{0, \pm 1, \ldots, \pm k\}$. NNHR is optimal in terms of max-min rates and the optimal rate is given by $R^*_{\text{e2e}} = \left(\frac{w_{0,0}}{R_a} + \frac{f(0,0) - w_{0,0}}{R_1}\right)^{-1}$. A simple hierarchical distributed scheduler that employs integrated access-backhaul, given in Algorithm 1, achieves the max-min optimal rate.*

*Proof.* See Appendix 5.9.1. □

Although the assumptions in Theorem 9 are idealistic, it gives an intuition that NNHR can be a good choice when the *bottleneck node* in the network is the fiber site and the effective load on the fiber site is well balanced in all four directions. A *bottleneck* node is formally defined as the node that has at least one link that is always active in order to attain the max-min rates. Also since the derived formula is simple, it offers a quick feasibility check for what is the maximum $k$ that supports a target rate. See Section 5.5.1 for a related discussion.

NNHR may not be desirable in all possible load conditions. However, since having dynamic routing requires exchange of control signals and a more complex system design, it would be desirable to design a system wherein some static routing always gives a reasonable performance. In order to do this network planning, which includes deciding how many antennas should be employed at different BSs in the $k-$ring deployment or their transmit powers, can play an important role. We now provide guidelines on network planning so that NNHR is justifiable in more general load scenarios.

If the BSs on the highways have much larger antenna gains than the non-highway relays then irrespective of the load it will be beneficial for the relays to employ the highway routing strategies since the highways links have much larger capacity to carry traffic than the non-highway links. We demonstrate this through a quick example. The relay on bottom left corner can trans-

**Algorithm 1** Theorem 9 scheduler

Denote $\mathcal{S}_r$ as the set of BSs in ring $r$, where $r = 0, \ldots, k$. Ring $r$ implying distance to $(0,0)$ is $r$D. Denote by $|\mathcal{S}_r|$ as the cardinality of the set. Total scheduling time is 1 unit.
**for** $r = 0 : k$ **do**
    **if** $r = 0$ **then**
        The MBS reserves $\frac{w_{0,0}\gamma^*}{R_a}$ fraction of time for access and rest for backhaul.
        The MBS equally divides the access (backhaul) time frame amongst respective users that need to be served over access (backhaul) links.
    **else**
        **for** $q = 1 : |\mathcal{S}_r|$ **do**
            Let $(i,j)$ be the BS indexed by $q$. The BS listens to its parent for backhaul for $\frac{f(i,j)\gamma^*}{R_1}$ units of time. This fraction is reserved by its parent already in previous for-loop iteration over $r$. Whenever the BS at $(i,j)$ is not listening, it reserves $\frac{w_{i,j}\gamma^*}{R_a}$ units of time for serving access and $\frac{\gamma^*(f(i,j)-w_{i,j})}{R_1}$ for transmitting on backhaul links. In the remaining time, which is non-negative, it stays silent.
            The BS equally divides the access (backhaul) time frame amongst respective users that need to be served over access (backhaul) links.
        **end for**
    **end if**
**end for**

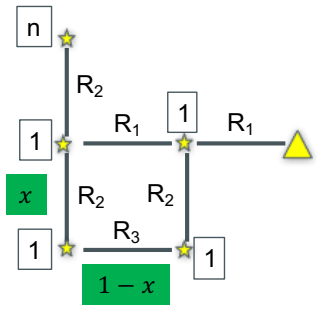



Figure 5.3: Justifying highway routing through an example.

mit towards the fiber site (denoted by the triangle) over two shortest nearest neighbour paths - transmitting to the relay above itself or transmitting to the relay on its right. To do load balancing, let it transmit $x$ fraction of its data to the relay above itself (route 1) and $1-x$ fraction of its data to the relay on the right (route 2). Let $\gamma w$ be the DL rate achieved by each relay in Fig. 5.3 with relay specific weights $w$ (can be interpreted as number of UEs) written in the square boxes, and we want to maximize $\gamma$. Since there are two routes from the relay to fiber site, we need a new definition for the e2e rate of the relay. The e2e rate is $\gamma w$ if the long term rate on each of the hops over route 1 is $x\gamma w$ and the long term rate on each of the hops over route 2 is $(1-x)\gamma w$. Similar to the proof of Theorem 9, the following set of inequalities need to be satisfied. $\gamma\left(\frac{n+4}{R_1}+\frac{n+1+x}{R_1}+\frac{1+1-x}{R_2}\right)\leq 1$, $\gamma\left(\frac{x}{R_2}+\frac{1-x}{R_3}\right)\leq 1$, $\gamma\left(\frac{n+1+x}{R_1}+\frac{n}{R_2}+\frac{x}{R_2}\right)\leq 1$, $\gamma\left(\frac{n}{R_2}\right)\leq 1$, $\gamma\left(\frac{1+1-x}{R_2}+\frac{1-x}{R_3}\right)\leq 1$, and $\gamma\left(\frac{n+4}{R_1}\right)\leq 1$. If $R_1$ and $R_2$ are large enough compared to $R_3$ for a fixed $n$, then the bottleneck inequality will be $\gamma < R_3/(1-x)$. The optimal choice for $x$ would be closer to 1 in order to maximize the upper bound on $\gamma$. As we have seen from Theorem 9, it is possi-

ble to construct a scheduling scheme that achieves the upper bound obtained from the inequalities written above. This motivates that when the relays on highways in $k-$ring deployment are made to have larger antenna gains than the non-highway relays, then the optimal routing paths will tend to be similar to NNHR strategies. We benchmark the performance of NNHR with theoretically optimal nearest neighbour routing (NNR) using the solution in [43] in the numerical results section for general load scenarios.

We now generalize the previous theorem for general load and unequal but deterministic instantaneous access rates to users. For simplicity of exposition, let us number the BSs in the network from 0 to $2k(k+1)$. BS index 0 corresponds to the MBS. With some abuse of notation $f(i)$ now denotes effective load on BS $i$ (number of UEs with the BS $i$ on it's route) with new indexing under some static routing strategy. Let us also number the users from 1 to $f(0)$ in ascending order of the index of their corresponding serving BS. $R_{a,u}$ denotes instantaneous access rates of users for $u = 1, \ldots, f(0)$.

**Theorem 10.** *Under a given nearest neighbour static routing strategy that determines the values of $f(i)$, $\forall i = 1, \ldots, 2k(k+1)$, $\gamma^* = \max_{\mathcal{S}} \min_u R_u^{\mathcal{S}} = \left(\max_{i \in \{0,\ldots,2k(k+1)\}} \mathbf{c}_i^T \mathbf{b}\right)^{-1}$, where $\mathcal{S}$ is restricted to all feasible scheduling matrices given the routing strategy,*

$$\mathbf{c}_i = \text{sum}\left(\mathbf{I}_{f(0)+1}, \sum_{r=0}^{i-1} w_r, \sum_{t=0}^{i} w_t\right) + (2f(i) - w_i)\, \mathbf{e}_{f(0)+1}, \forall i \neq 0,$$

$$\mathbf{c}_0 = \text{sum}\left(\mathbf{I}_{f(0)+1}, 0, w_0\right) + (f(0) - w_0)\, \mathbf{e}_{f(0)+1},$$

*where* $\mathbf{e}_j$ *represents the* $j^{th}$ *column of identity matrix of dimension* $f(0)+1$ *and* $\text{sum}\left(\mathbf{I}_{f(0)+1}, l, u\right) = \sum_{j=l+1}^{u} \mathbf{e}_j$. *Here,* $\mathbf{b} = \left[\frac{1}{R_{a,0}} \ \frac{1}{R_{a,1}} \ \ldots \ \frac{1}{R_{a,f(0)}} \ \frac{1}{R_1}\right]^T$, *where* $R_{a,i}$ *is the access rate to* $i^{th}$ *user.*

*Proof.* If $\gamma$ is a minimum rate achieved by all users, then for BS with index $i$ the following inequality should be satisfied.

$$\gamma\left(\sum_{t=l_i+1}^{l_i+w_i}\frac{1}{R_{a,t}} + \frac{f(i)-w_i}{R_1} + \frac{\mathbb{1}(i\neq 0)f(i)}{R_1}\right) \leq 1, \forall i = 0,\ldots,2k(k+1),$$

where $l_i = \sum_{r=0}^{i-1} w_r$ ensures that the indices from $l_i+1$ to $l_i+w_i$ correspond to UEs associated with BS $i$. This inequality can be written as $\gamma \mathbf{c}_i^T \mathbf{b} \leq 1$. To justify that the upper bound on max-min rate given by $\gamma^* = \frac{1}{\max_i \mathbf{c}_i^T \mathbf{b}}$, the following scheduler is sufficient. The MBS first allocates $\frac{(f(0)-w_0)\gamma^*}{R_1}$ fraction of time for backhaul and equally divides the time amongst the $f(0)-w_0$ users which are eventually served over the backhaul links connected to the MBS. The MBS allocates $\gamma^*/R_{a,t}$ fraction of time for user $t$ connected to MBS, where $t = 1,\ldots,w_0$. Then in the time when a relay in ring 1 is not scheduled by the MBS, it allocates $\frac{(f(i)-w_i)\gamma^*}{R_1}$ fraction of time for serving over backhaul links away from the fiber site, which is equally divided for transmitting data of each of the $f(i)-w_i$ users. The relay allocates $\gamma^*/R_{a,t}$ fraction of time for user $t$ connected to it, where $t = 1,\ldots,w_i$. This process continues hierarchically for all relays in the $k-$ring deployment. □

Now, we extend the result in Theorem 10 to full duplex BSs. Instead of the required fraction of time for reception plus transmission being $\leq 1$ for

every BS, we now have two separate inequalities per BS – one for transmission time and one for reception time. Since there will be self-interference at each relay the access rates and backhaul rates will be different than in Theorem 10. Let access rates to users under full duplex relaying be $R^f_{a,i} \leq R_{a,i}$ and the single hop backhaul rate be $R^f_1 \leq R_1$. Although the system model set up in Section 5.3 was for DL, all the relevant definitions can be extended for UL as well. We now consider a scenario when there are some UL and some DL UEs. Note that a UE cannot be both UL and DL. Let $\mathcal{U}_{\mathrm{DL}}$ and $\mathcal{U}_{\mathrm{UL}}$ be the set of indices of downlink (DL) and uplink (UL) UEs. Let $w_i^{\mathrm{DL}}$ and $w_i^{\mathrm{UL}}$ denote the number of DL and UL UEs connected to BS $i$. Similarly, $f^{\mathrm{DL}}(i)$ and $f^{\mathrm{UL}}(i)$ corresponds to effective DL and UL load on BS indexed by $i$.

**Theorem 11.** *Considering full duplex BSs, and under a given nearest neighbour static routing strategy that determines the values of* $f^{\mathrm{DL}}(i)$ *and* $f^{\mathrm{UL}}(i)$, $\forall i = 1, \ldots, 2k(k+1)$, $\gamma^* = \max_{\mathcal{S}} \min_u R_u^{\mathcal{S}} = \min(\gamma_{\mathrm{tx}}, \gamma_{\mathrm{rx}})$, *where* $\mathcal{S}$ *is restricted to all feasible scheduling matrices given the routing strategy,* $\gamma_{\mathrm{tx}} = \left(\max_{i \in \{0,\ldots,2k(k+1)\}} \mathbf{c}_{\mathrm{tx},i}^T \mathbf{b}_f\right)^{-1}$ *and* $\gamma_{\mathrm{rx}} = \left(\max_{i \in \{0,\ldots,2k(k+1)\}} \mathbf{c}_{\mathrm{rx},i}^T \mathbf{b}_f\right)^{-1}$. *Here,* $\mathbf{b}_f = \begin{bmatrix} \frac{1}{R^f_{a,0}} & \frac{1}{R^f_{a,1}} & \cdots & \frac{1}{R^f_{a,f(0)}} & \frac{1}{R^f_1} \end{bmatrix}^T$,

$$\mathbf{c}_{\mathrm{tx},i} = \mathrm{sum}^{\mathrm{DL}}\left(\mathbf{I}_{f(0)+1}, \sum_{k=0}^{i-1} w_k, \sum_{k=0}^{i} w_k\right) + \left(f(i) - w_i^{\mathrm{DL}}\right) \mathbf{e}_{f(0)+1}, \forall i \neq 0,$$

$$\mathbf{c}_{\mathrm{tx},0} = \mathrm{sum}^{\mathrm{DL}}\left(\mathbf{I}_{f(0)+1}, 0, w_0\right) + \left(f^{\mathrm{DL}}(0) - w_0^{\mathrm{DL}}\right) \mathbf{e}_{f(0)+1},$$

$$\mathrm{sum}^{\mathrm{DL}}(\mathbf{I}_{f(0)+1}, l, u) = \sum_{j=l+1}^{u} \mathbf{e}_j \mathbb{1}\left(\text{UE } j \text{ is DL}\right),$$

*where* $\mathbf{e}_j$ *represents the* $j^{th}$ *column of identity matrix of dimension* $f(0)+1$. $\mathbf{c}_{\mathrm{rx},i}$ *is same as* $\mathbf{c}_{\mathrm{tx},i}$ *but with superscript* DL *replaced by* UL *in all places.*

The proof of Theorem 11 is similar to that of Theorem 10. With these general formulae for IAB, we next turn our attention to the analysis of a couple of OAB schemes.

### 5.4.2 Orthogonal access backhaul

Max-min optimization with IAB may face difficulties for practical implementation, owing to the need to know global information of load and access rates for solving the optimization problem. The OAB schemes discussed here are potentially simpler to implement. Let $\zeta$ be the fraction of resources reserved for access and rest are reserved for backhaul. We now perform optimization only over entries of the scheduling matrices for backhaul links. Every BS is assumed to divide the access time equally amongst all UEs directly associated with it. Furthermore, now all UEs associated with a BS have same route from the fiber site to the associated BS.

If a backhaul link with instantaneous rate $R$ is activated for $\tau$ fraction of time to serve all UEs associated with a relay, then *long term backhaul rate of a relay on a link* is defined as $\tau R$. Furthermore, e2e backhaul rate of a relay is defined as minimum of long term backhaul rate of the relay over each hop from the fiber site to the relay.

We first consider a simple OAB scheme wherein equal e2e rate is offered to each relay. As this scheme does not optimize the rates based on load per

BS, there will be some over-utilized and some under-utilized BSs. This issue, however, can be addressed by enabling dual connectivity which we will study in the next section. For simplicity of exposition, assume DL backhauling. The analysis holds for a mix of DL and UL backhauling since instantaneous link rates on backhaul do not change for UL and DL.

**Theorem 12.** *If all relays are offered an equal e2e backhaul rate, say $\gamma$. Maximizing $\gamma$ over all possible backhaul schedules, we get optimal $\gamma^* = \frac{(1-\zeta)R_1}{2k(k+1)}$ and NNHR is optimal. Furthermore, e2e rate for any user connected to some BS at $(i,j)$ is given by $\frac{1}{w_{i,j}} \min(\zeta R_a, \frac{(1-\zeta)R_1}{2k(k+1)})$, assuming every BS divides the access and backhaul time equally amongst the users directly associated with that BS.*

*Proof.* Let $\gamma$ be the maximum long term rate offered to each relay assuming NNHR. Then the following should be satisfied. $\gamma/R_1$ is the minimum fraction of resources that are allocated for backhauling to each of the $2k(k+1)$ relays by the MBS. Thus, $\gamma(2k(k+1))/R_1 \leq (1-\zeta)$. Let $f(i,j)-1$ represent total number of relays served by $(i,j)$. The following inequalities should also hold. $\gamma\left(\frac{f(i,j)-1}{R_1} + \frac{f(i,j)}{R_1}\right) \leq 1-\zeta$, for all $(i,j) \neq (0,0)$. Here, $\gamma f(i,j)/R_1$ is the fraction of time for relaying data to $(i,j)$ from the fiber site. $\frac{f(i,j)-1}{R_1}$ is the fraction of time for relaying data from $(i,j)$ to the BSs away from the fiber site. Since $2f(i,j)-1 < 2k(k+1)$, which holds because $f(i,j) = f(-i,-j)$ considering NNHR, the least upper bound on $\gamma$ is $\gamma \leq (1-\zeta)R_1/2k(k+1)$. This is achieved by using a scheduler similar to Algorithm 1. The main difference is that $R_a$ is set to $\infty$, which makes time allocated for access equal to zero,

and $w_{i,j} = 1$ making $f(i,j)$ as the effective number of relays served by $(i,j)$ including itself.

A non-NNHR scheme cannot offer rates higher than $(1-\zeta)R_1/2k(k+1)$ as the fiber site will always have to support at least $2k(k+1)$ relays, irrespective of the routing scheme. This proves $\gamma = (1-\zeta)R_1/2k(k+1)$. By definition, the e2e rate for a user is the minimum of its access long term rate and e2e backhaul rate. Consider a UE connected to $(i,j)$. Since backhaul rate to $(i,j)$ is equally divided amongst all $w_{i,j}$ users, the e2e backhaul rate of the UE is $\frac{(1-\zeta)R_1}{w_{i,j}2k(k+1)}$. Long term access rate of the UE is $\zeta R_a/w_{i,j}$, since each user connected to a relay receives equal fraction of time for access. Thus, the e2e rate for the user is given by $\frac{1}{w_{i,j}} \min(\zeta R_a, \frac{(1-\zeta)R_1}{2k(k+1)})$. □

**Corollary 6.** *If $w_{i,j} = w_{-i,-j}$ and $w_{0,0} \geq w_{i,j}$ and access rates to all UEs are given by $R_a$, there exists an OAB strategy that performs as good as IAB in terms of max-min rates.*

*Proof.* Consider the following OAB scheme. $\zeta$ is the fraction of access resources (also called as slots) and $1-\zeta$ is the fraction of backhaul slots. Within the backhaul slots, target long term rate to each relay is $\gamma w_{i,j}$, for all $i,j$. Resources are allocated to maximize $\gamma$.

Assuming NNHR, similar inequalities as (5.1) and (5.2) can be written to find an upper bound on $\gamma$ with the following differences. Since access resources are orthogonal from backhaul resources, the terms of the form $w_{i,j}/R_a$ are not present but the rest of the terms remain the same since e2e backhaul

rate per relay is $\gamma w_{i,j}$. Thus, it is easy to see that $\gamma \leq \frac{(1-\zeta)R_1}{f(0,0)-w_{0,0}}$. Achieving the upper bound is possible employing a scheduler same as Algorithm 1 with the following difference. Set $R_a \to \infty$ to make sure there are no access slots allocated in the backhaul resource blocks. Optimality of NNHR is argued exactly as in proof of Theorem 9. This implies maximum achievable $\gamma = \frac{(1-\zeta)R_1}{f(0,0)-w_{0,0}}$.

Thus, with the OAB scheme under consideration the e2e rate for a user connected to a BS at $(i,j)$ is given by $\min(\frac{\zeta R_a}{w_{i,j}}, \frac{(1-\zeta)R_1}{f(0,0)-w_{0,0}})$, assuming round robin scheduling done by $(i,j)$ amongst $w_{i,j}$ UEs for access and that the e2e backhaul rate to $(i,j)$ was equally divided amongst all $w_{i,j}$ UEs. Minimum e2e rate corresponds to $i=j=0$. Maximizing minimum e2e rate over $\zeta$, it is found that the max-min rate equals $\left(\frac{w_{0,0}}{R_a} + \frac{f(0,0)-w_{0,0}}{R_1}\right)^{-1}$, same as Theorem 9. $\square$

End-to-end backhaul rate with the OAB scheme described in the proof of Corollary 6 can be analyzed in general load scenario, like in Theorem 10. In this case, the e2e backhaul rate is exactly same as that in Theorem 10 but replacing $1/R_{a,u}$ by 0 in the definition of vector $\mathbf{b}$. If long term rate of a UE on a link is defined as $\lim_{t\to\infty} \frac{1}{t}\int_0^t X(\tau)\mathrm{d}\tau$, where $X(\tau)$ denotes a stationary ergodic random process and denotes the instantaneous access rate of the user as a function of time, then using ergodic theorem the e2e rates in Theorem 12 can be extended even when access rates are random variables by replacing $R_a$ with $\mathbb{E}[R_a]$.

## 5.5 Example Applications of the Analysis

In this section, we discuss some simple use-cases of our analysis.

### 5.5.1 5G Networks with Minimum Rate of 100 Mbps

Deploying a new cellular network operating at mmWave involves significant cost and time overheads. Thus, it does not make sense if the deployed mmWave network offers only marginal gains over existing 4G networks. A minimum 100 Mbps per UE target has been set for 5G networks operating at mmWave frequencies. The analysis can be used to evaluate feasibility of potential BS or UE deployments for 5G networks.

#### 5.5.1.1 Minimum number of rings required to get 100 Mbps rates

A closed-form expression for maximum $k$ that supports 100 Mbps per UE can be obtained in simple settings like Theorem 9.

**Corollary 7.** *If all relays have equal load $w$, then the maximum $k$ that can still meet the max-min target rate of $\gamma_{\text{target}}$ is given by $k \leq \frac{\sqrt{1+2R_1\left(\frac{1}{w\gamma_{\text{target}}}-\frac{1}{R_a}\right)}-1}{2}$. There exists no solution if $\gamma_{\text{target}} > \frac{R_a}{w}$.*

*Proof.* The max-min rate is given by $\gamma^* = \frac{1}{w}\left(\frac{1}{R_a} + \frac{2k(k+1)}{R_1}\right)^{-1}$. Rearranging and solving the quadratic equation we get the result by using $\gamma^* \geq \gamma_{\text{target}}$. □

Example. If $\gamma_{\text{target}} = 100$ Mbps and $w = 5$ full buffered active UE per BS, access rates should be equal to at least 500 Mbps to meet this for $k = 0$.

Now suppose $R_1 = 10$ Gbps and we need to design a network with $k = 2$ rings. The required access rate is $R_a \geq 2.5$ Gbps. Thus, with a bandwidth of 1 GHz this translates to a spectral efficiency of at least 10 bps/Hz for backhaul links and a spectral efficiency of 2.5 bps/Hz on access links. Using the physical layer models in Remark 8 and Remark 9, the following configuration can meet this requirement considering worst case NLOS UEs on the street at a distance of 50m from the BSs, which are spaced on a grid with ISD= 100m . BSs have 64 antennas, UEs have 16 antennas, and transmit power is 1W for all devices. These numbers are reasonable; 5G mmWave access points will have up to 1024 antennas, with UEs having up to 64 antennas [16].

#### 5.5.1.2 Soft max-min

Strictly maximizing the minimum rate in a mmWave system may lead to very poor e2e rates achieved by all UEs if a few of the UEs have very poor spectral efficiency, e.g. they are severely blocked by surrounding objects. Thus, it is practically beneficial to softly optimize the max-min rates. Here, we demonstrate a possible procedure. UEs that have very poor spectral efficiency, denoted as "bad UEs", are placed with pseudo UEs for finding max-min rates. The pseudo UEs fake a higher SINR for the corresponding "bad UEs". This allows the rest of the "good UEs" to have much better rates after max-min optimization is performed. Essentially, these "bad UEs" sacrifice themselves for the benefit of the whole. In a practical 5G system, such UEs would soon switch to a sub-6GHz legacy band to maintain a minimum performance level.

### 5.5.2 Analyzing performance of dual-connectivity.

Multi-connectivity, wherein a UE connects to multiple BSs on the same or different bands, can counteract dynamic blocking in mmWave cellular [136]. For self-backhauled networks, dual connectivity has another advantage, which is smoothing out the load imbalance across relays. This will make resource allocation simpler in self-backhauled networks since employing equal rate per relay OAB is much simpler than IAB. Here, we look at a specific implementation of dual-connectivity. OAB is assumed with $\zeta$ fraction of resources for access. Optimization is done to offer equal backhaul rates per relay. Consider a user connected to two BSs offering least path loss. Consider a user connected to relays at $(i, j)$ and $(i-1, j)$. Let the distance from the two BSs be $x$ and $y(< x)$, respectively. It is assumed that the user has at least two RF chains so that it can receive signals from both connected BSs simultaneously. $R_a(x)$ is the access rate to the user from BS at $(i, j)$ and $R_a(y)$ is the access rate from BS at $(i-1, j)$. Let $R_{single}$ and $R_{dual}$ be the rates of the user under single and dual connectivity. Using Theorem 12,

$$R_{single} = \frac{1}{w_{i,j}} \min\left(\zeta R_a(x), \frac{(1-\zeta)R_1}{2k(k+1)}\right)$$

and

$$R_{dual} = \frac{1}{w'_{i,j}} \min\left(\zeta R_a(x), \frac{(1-\zeta)R_1}{2k(k+1)}\right) + \frac{1}{w'_{i-1,j}} \min\left(\zeta R_a(y), \frac{(1-\zeta)R_1}{2k(k+1)}\right),$$

where $w'_{i,j}$ $(\geq w_{i,j})$ is the new load on $(i, j)$ after dual connectivity.

The above formula even works in scenarios when there is dynamic blocking, in which case $R_a(x)$ and $R_a(y)$ are long term access rates, that is averaged over the toggling between LOS and NLOS state of the service links.

**Remark 8** (Computing access rates)**.** $R_a(x) = \mathrm{W} \min\left(\log_2\left(1 + \mathrm{SNR}_a\right), \mathrm{SE}_{\max}\right)$, *where* $\mathrm{SNR}_a$ *is the effective received signal power to noise power ratio and is given by*

$$\mathrm{SNR}_a = \left(\frac{\sigma^2}{P_r} + \frac{1}{\mathrm{SNR}_{\max} N_r}\right)^{-1}.$$

*Here,* $P_r/\sigma^2$ *is the actual signal to noise ratio (SNR) as defined next, and* $\mathrm{SNR}_{\max} N_r$ *limits the maximum possible received SNR with* $N_r$ *equal to the number of receiver antennas. A similar model for dampening very high SNR due to device imperfections is common in the industry, e.g. see the Qualcomm paper [137]. It can be derived by modeling a virtual amplify-and-forward transmission hop within the receiving device, which leads to effective SNR being half of the harmonic mean of the actual and maximum SNR [138, (4)]. Note that for large* $\mathrm{SNR}_{\max} N_r$*, the effective SNR is close to* $P_r/\sigma^2$*. However, if* $P_r/\sigma^2$ *is itself very large, then the SNR cannot exceed* $\mathrm{SNR}_{\max} N_r$*. Note that* $\mathrm{SE}_{\max}$ *is the limit on maximum spectral efficiency, which is related to modulation and coding employed by the receiver. Here,* $P_r = \left(\frac{\lambda}{4\pi}\right)^2 \Upsilon \mathrm{P} \mathrm{N}_{\mathrm{BS}} \mathrm{N}_{\mathrm{UE}} x^{-\alpha}$*, where* P *is the transmit power,* $\sigma^2$ *is the noise power,* W *is the bandwidth,* $\mathrm{N}_{\mathrm{BS}}$ *and* $\mathrm{N}_{\mathrm{UE}}$ *are the number of antennas at the BS and UE,* $\lambda$ *is the wavelength in meters,* $\Upsilon$ *is the blockage dependent correction factor [2], and* $\alpha$ *is the blockage dependent path loss exponent (PLE). If the link is LOS, then* $\alpha = \alpha_l$ *and* $\Upsilon = 1$*. If the link is NLOS, then* $\alpha = \alpha_n$ *and* $\Upsilon = \Upsilon_n \ll 1$*.*

Table 5.1: Default numerical parameters

| **Notation** | **Parameter(s)** | **Value(s) if applicable** | **Notation** | **Parameter(s)** | **Value(s) if applicable** |
|---|---|---|---|---|---|
| $f_c$ | Carrier frequency | 28 GHz [33] | W | Total bandwidth | 800 MHz [33] |
| $P_d$ | BS transmit power | 30 dBm [33] | $P_u$ | UE transmit power | 23 dBm [33] |
| $\eta$ | Fraction DL UEs | 1 | $\sigma^2$ | Noise power | $-174 + 10\log_{10}(\mathrm{W}) + 10$ dBm |
| $\alpha_l$ | LOS PLE | 2 [24] | $\alpha_n$ | NLOS PLE | 3.4 [24] |
| $\mathrm{N_{BS}}$ | BS antennas | 64 [16] | $\mathrm{N_{UE}}$ | UE antennas | 16 [16] |
| D | ISD | 200m | $k$ | Number of rings | 3 |
| $\Upsilon_n$ | Correction factor | $-5$dB [2, 42] | $\mathrm{SE_{max}}$ | Maximum spectral efficiency | 10 bps/Hz [139] |

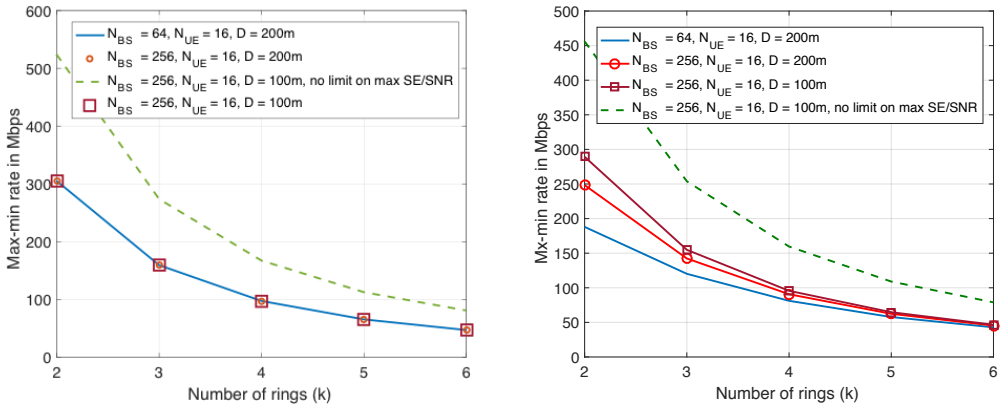
**Remark 9** (Computing backhaul rate)**.** $R_1 = \mathrm{W}\min\left(\log_2\left(1+\mathrm{SNR}_b\right), \mathrm{SE_{max}}\right)$, *where* $\mathrm{SNR}_b$ *is half of the harmonic mean of* $\frac{(\lambda/4\pi)^2 \mathrm{P N_{BS}^2 D^{-\alpha_l}}}{\sigma^2}$ *and* $\mathrm{SNR_{max} N_{BS}}$.

## 5.6 Numerical Results and Design Guidelines Based on Analysis

In this section, we evaluate the derived formulae to explore system design insights for multi-hop mmWave cellular networks. In the next section, the main analytical assumption – noise-limitedness – will be validated. Table 5.1 summarizes key parameters which are fixed throughout the numerical study, unless specified otherwise. NNHR is assumed, unless specified otherwise. We

choose $\text{SNR}_{\text{max}} = 16$ dB, so that the maximum received SNR at UEs equals 28 dB considering 16 antennas, which is close to the 30dB value in [137]. For backhaul links, the maximum received SNR is 34 dB considering 64 antennas. For 5G-NR, it is possible to support up to 1024 QAM [139] and thus $\text{SE}_{\text{max}} = 10$ bps/Hz is chosen.

**Fall in throughput with number of rings.** To understand the fall in throughput with number of rings, we consider 2 worst case UEs per BS located at a distance D/2 on the streets. LOS access and all DL UEs is assumed. Fig. 5.4(a) shows the fall in throughput with number of rings. It is surprising to note that it is possible to achieve minimum 100 Mbps per UE with even a 4 ring deployment, which covers an area of $800 \times 800$ m$^2$ and supports 40 relays per fiber site. Having a larger $\text{N}_{\text{BS}}$ hardly changes the rate as the network is backhaul limited with backhaul links operating at $\text{SE}_{\text{max}}$. Decreasing D to 100 meters also does not change the rates. As per Corollary 7, throughput decays as $\frac{1}{w}\left(\frac{1}{R_a} + \frac{2k(k+1)}{R_1}\right)^{-1}$. Since we consider LOS UEs, $\text{R}_a$ is already saturated by $\text{SNR}_{\text{max}}$ for D = 200m. Also, $R_b$ is limited by $\text{SE}_{\text{max}}$ and does not change by decreasing D. However, note that 2 UEs per BS with D = 100m itself supports 4x higher user density than for D = 200m. If there were no limit on spectral efficiency or SNR, then even up to $k = 6$, that covers an area of $1.2 \times 1.2$ km$^2$, can be supported with user density of 200 UEs/km$^2$. This result motivates supporting higher order modulations and high SNRs for enabling large scale mmWave mesh network deployments for 5G.




(a) 2 LOS UEs per BS located at 100 m from the serving BS.

(b) 2 NLOS UEs per BS located at 100 m from the serving BS.

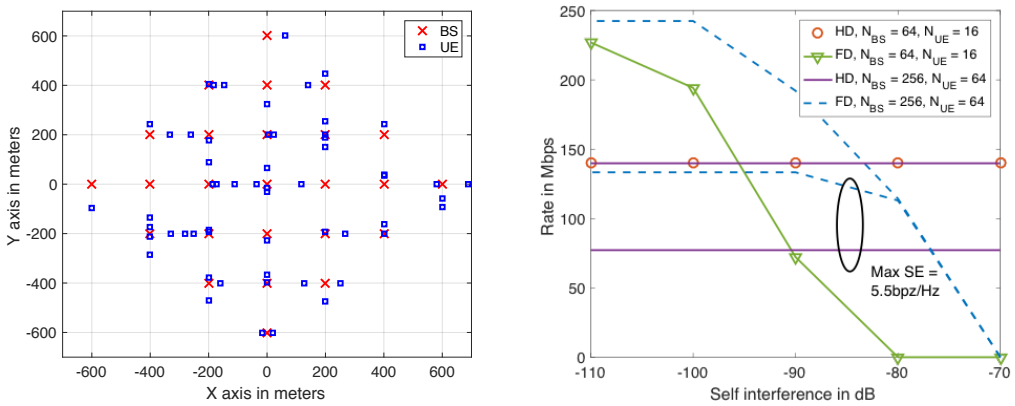
Figure 5.4: Fall in throughput with $k$.

As can be seen in Fig. 5.4(a), throughput decays quickly with $k$ as the networks are backhaul limited. For large values of $k$, when the $1/R_a$ term is negligible, throughput decays by a factor of $k/(k+2)$ as $k$ increments by 1. The $1/R_a$ factor makes throughput decay slightly slower than above for smaller values of $k$. More specifically, if one fits function $\alpha/k^{\beta}$ to the plot for $\mathrm{N_{BS}} = 64$ and $\mathrm{N_{UE}} = 16$, then $\beta = 1.6$. The decay is slower in access limited networks, when $1/R_a$ term is non-negligible. This can be observed from Fig. 5.4(b), which reproduces the scenarios in Fig. 5.4(a) but with NLOS UEs. Note that up to 3 rings can be supported even with NLOS UEs.

We now consider a more general UE deployment setup as shown in Fig. 5.5(a). On an average there are 2 UEs per BS in the 3-ring deployment. A random realization of LOS/NLOS states for UE to/from BS links was generated considering 50% probability of being LOS within a distance of 200m.

Minimum path loss association is done. For the realization considered, 55% UEs connected to LOS BSs. Also by default $\eta = 0.5$, that is about 50% UEs are DL and rest are UL. Spectral efficiency (SE) has a minimum limit of 0.02 bps/Hz below which the rate is zero.

**Impact of Full Duplex Relays.** Fig. 5.5(b) shows the comparison of full and half duplex relaying. X axis is the self-interference (SI) introduced by full duplexing and Y axis plots the optimal rates in Mbps. We consider soft max-min optimization, introduced in Section 5.5.1, wherein 10% of bad UEs are replaced with pseudo UEs that fake an arbitrarily large rate. We consider soft max-min since we observe that considering max-min optimization in the considered setup leads to a conclusion that full duplexing can provide higher rates than half duplex only if SI is less than $-110$dB, which is impractical to achieve as per state of the art prototypes [140]. Fig. 5.5(b) explores scenarios wherein larger SI can be tolerated. Even with soft max-min optimization, significant gains with full duplexing are observed for the default setup only if SI$< -100$dB. Fig. 5.5(b) shows that considering larger antenna gains at the BSs and UEs helps increase the requirement of maximum tolerated SI to $-90$dB. Considering a maximum spectral efficiency of 5.5 bps/Hz further increases the tolerance of SI to $-80$dB, which is practical [140]. Note that 5.5bps/Hz corresponds to spectral efficiency with 64 QAM and light coding. Similar values of $\text{SE}_{\text{max}}$ have been used significantly in prior work [24, 141]. We next turn our attention to understanding if OAB can closely follow the rates obtained using IAB.




(a) Topology under consideration.

(b) Full versus half duplex relaying with soft max-min.

Figure 5.5: Impact of full duplex relaying on max-min rates.

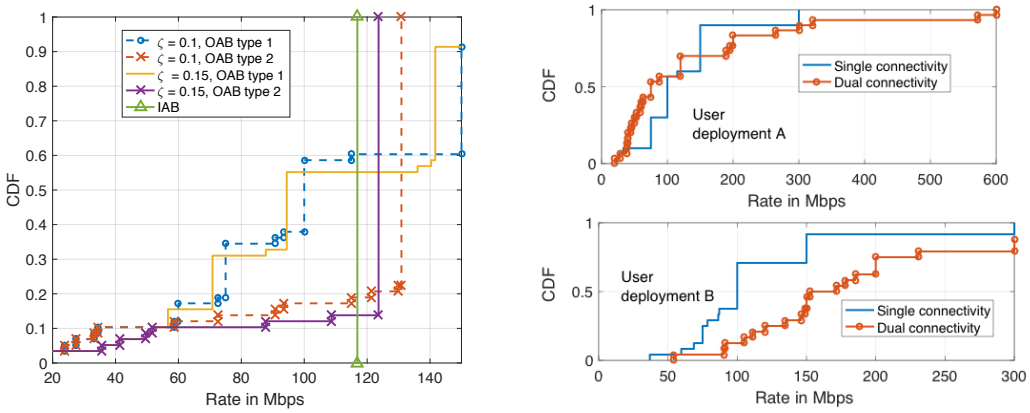
**OAB versus IAB.** The distribution of e2e rates obtained using OAB is compared with IAB in Fig. 5.6(a). We consider two types of OAB. First allocates equal backhaul rate to each relay (called type 1). Second type offers a backhaul rate $w_{i,j}\gamma$ to a relay at $(i, j)$, wherein maximum $\gamma$ is computed (called type 2). The max-min rates with IAB outperform the rate obtained by more than 60% of UEs with OAB type 1. Although not shown in the plot, varying $\zeta \in (0, 1)$ does not change this insight. However, it is interesting to note that with OAB type 2 it is possible to achieve rates slightly greater than IAB rates for about 85% UEs by choosing $\zeta = 0.15$. This is encouraging for practical implementations since OAB type 2 requires less global information for performing the optimization as compared to IAB.

**Dual connectivity versus single connectivity.** We conclude our discussion of design insights based on the analysis by evaluating the benefit

of dual-connectivity as described in Section 5.5.2. Fig. 5.6(b) plots the rates with single and dual connectivity (SC and DC) considering two deployments and OAB type 1. Deployment A is the one in Fig. 5.5(a), wherein there are about 2 UEs per BS on an average with a load variance of 1.1. Deployment B is not shown for space constraints and has same mean UEs per BS but variance is 2.3. For deployment B, median rates with DC are almost 1.5x higher than SC. Although the load per BS potentially becomes higher with DC, the load imbalance across BSs is reduced. Since equal backhaul rate per relay is offered, load balancing makes it possible to exploit the backhaul links which were underutilized using SC. However, note that the rates with DC are roughly similar to SC for deployment A with lower load imbalance of UEs across BSs. We observe that in general the higher the load imbalance with SC, the higher the gain in data rates with DC. Up to 2x gains with DC were observed in UE deployment scenarios with all LOS access links.

## 5.7 Validation of Noise-limitedness.

Same default parameters as Table 5.1 are used in this section. The goal is to motivate why noise-limited analysis works through a couple of empirical observations. Also, we observe NNHR operates optimally even in more general scenarios than in Theorem 9. We also propose a greedy variant of PF scheduling for multi-hop networks in one of the numerical examples that is used to validate noise-limited analysis. This example is also useful to show how the analysis can be used as a benchmarking tool for complex simulators. All UEs




(a) OAB type 2 closely follows IAB. (b) Impact of dual connectivity on rate.

Figure 5.6: OAB vs IAB, and impact of dual connectivity.

are assumed to be DL.

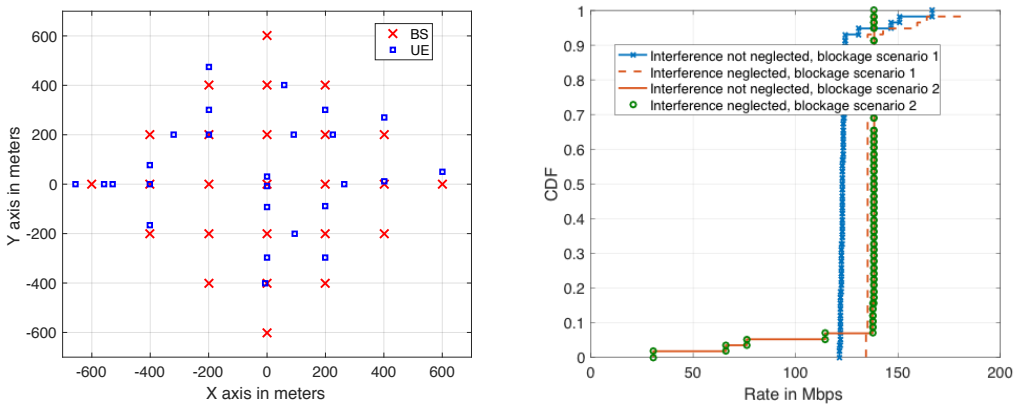
### 5.7.1 Few bottleneck links helps noise-limitedness.

We compare the max-min rate obtained from our noise-limited analysis with that computed using the linear programming (LP) solution in [43], which jointly optimizes the scheduling and routing. An arbitrary deployment was considered and interference was not neglected in [43]. This, however, lead to a LP formulation with very high numerical complexity as compared to our work. Specifically, if there are $L$ links in the network one needs to create matrices of the size on the order of $2^L$ to implement the LP. Although [43] discusses some methods to reduce this complexity, the exact quantification is unknown.

**Simulation setup.** We consider the deployments in Fig. 5.5(a) and Fig. 5.7(a) with average loads equal to 4 UEs per BS and 2 UEs per BS. All

access links are LOS and nearest neighbour BS association is done. All UEs are DL. Searching over all possible routes is not possible using the algorithm in [43] considering that the network under our consideration has 25 BSs and 99 UEs. We reduce the search space by considering only NNR (not necessarily highway routing) on the grid. Since links across orthogonal streets can have very high path loss [114] and long links on the same street would tend to be NLOS, this is likely not a bad assumption for mmWave mesh networks. Also since listing all scheduling patterns given NNR is itself time and memory intensive (since there are 135 valid links in Fig. 5.5(a) for example even after reducing the search space for routing and there will be on the order of $2^{135}$ potential schedules), we do a greedy search to list transmission schedules. Banking on the possible noise-limitedness, we first greedily list 50 transmission schedules that pack as many links together as possible still respecting the half duplex constraint, half of them forced to have at least one backhaul link connected to the fiber site. Then we append all transmissions schedules which have 3 links active at a time to make sure the LP has a solution.

To model interference, we consider received signal power from interfering transmitters as follows. $P_r = (\lambda/4\pi)^2 \, \mathrm{P} G_t G_r x^{-\alpha}$, where $G_t$ and $G_r$ are random antenna gains and rest of the parameters are defined in Remark 8. The antenna gains equal to $G_{\max}$ if the interfering link is pointed at the receiver (note that we have only 4 directions to point in the grid deployment) and $G_{\min}$, otherwise. Here, $G_{\max} \in \{\mathrm{N_{BS}}, \mathrm{N_{UE}}\}$ depending on whether the transmitter or receiver is a BS or UE, and $G_{\min}(\mathrm{dB}) = G_{\max}(\mathrm{dB}) - 25\mathrm{dB}$,




(a) Validation topology.

(b) Performance of proposed greedy PF scheduler.

Figure 5.7: Validation plots.

where 25 dB is the front to back ratio. Note that NLOS interferers never have their beams aligned in our interference model and this 25 dB loss in antenna gain can be reinterpreted as the corner loss considering the NLOS beams are always aligned [114, 134]. It was shown in [134] that NLOS interferers contribute negligibly to total interference in an urban canyon type deployment model.

**Comparing max-min rates with and without interference.** Table 5.2 summarizes the results of max-min rates obtained by running the LP in [43] for different scenarios. Specifically, max-min rates were computed assuming noise-limitedness and considering interference. Also, max-min rates were computed assuming NNHR and without such an assumption. Here, $t_1$ denotes the fraction of time that the greedy schedules were used, so $1 - t_1$ is the fraction of time 3 links were active at a time. The fraction of time wherein

Table 5.2: Empirical evidence for noise-limitedness and justification of highway routing.

| **Scenario (Fig. 5.5(a))** | $t_1$ | $t_2$ | **Max-min rate (Mbps)** |
|---|---|---|---|
| Optimal NNR + interference | 0.03 | 1 | 137.15 |
| Optimal NNR without interference | 0.04 | 1 | 137.17 |
| NNHR + interference | 0.04 | 1 | 137.08 |
| NNHR without interference | 0.03 | 1 | 137.17 |
| **Scenario (Fig. 5.7(a))** | $t_1$ | $t_2$ | **Max-min rate (Mbps)** |
| Optimal NNR + interference | 0.00 | 1 | 304.64 |
| Optimal NNR without interference | 0.01 | 1 | 304.64 |
| NNHR + interference | 0.01 | 1 | 304.64 |
| NNHR without interference | 0.01 | 1 | 304.64 |

at least one link connected to the fiber site was active is denoted by $t_2$.

It is surprising to note that irrespective of whether interference is considered or not, the max-min rates do not change. Furthermore, the rates do not change irrespective of whether NNHR or optimal NNR is considered. Also note that the rate corresponding to Fig. 5.5(a) is exactly same as the IAB rate in Fig. 5.6(a) using our analysis. Similarly it was confirmed that the max-min rate corresponding to Fig. 5.7(a) is equal to that from our analytical result in Theorem 10. These observations are explained by noting the values of $t_1$. Since $t_1 \ll 1$, most of the schedules used only 3 active links at a time to meet the max-min rates. In other words, the optimal scheduler did not use the transmission schedules with greedy packing. Thus, the interference is negligible in these scenarios. As mentioned in [43] there is no unique solution to the LP and thus the values of $t_1$ are not unique. The key takeaway, however, is that there exists a solution that achieves max-min rates under the scheduling

and routing search space considered with small $t_1$. Since $t_2 = 1$, it implies the bottleneck node is the fiber site. This exercise highlights the importance of our noise-limited analysis and also makes a case for near optimality of NNHR.

Similar observation highlighting noise-limitedness due to interference aware schedulers were reported in [40, 43]. If one comes across a deployment and traffic scenario wherein the rates with NNHR are much lower than optimal NNR, then the methodology discussed in Section 5.4.1 to increase the antenna gains on highway relays to make static highway routing still reasonable can be attempted. Identifying such scenarios and evaluating how large antenna gains on highway relays should be compared to non-highway relays is a scope for future work. Our code for implementing the LP in [43] is available at [142].

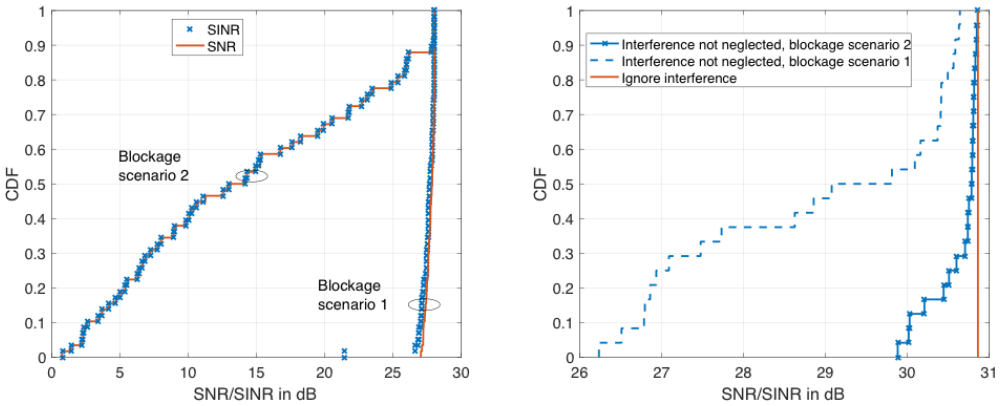
**Remark 10.** *Since our analytical results with NNHR give the same rate, the results for NNHR in Table 5.2 are accurate in spite of a small search space. Furthermore, increasing the number of greedy schedules to 1800 did not change the result for the case of NNR without interference, making us confident on the near optimality of NNHR in the scenario considered.*

Noise-limitedness is observed in the validation results since the scheduler is intelligent to pick the right balance of choosing different combination of 3 active links at a time to avoid interference on the bottleneck links. Developing practical schedulers which identify and protect the bottleneck links from interference is one avenue of further research. We now report another interesting observation that shows that even if schedulers are not interference-aware,

noise-limited analysis can still provide accurate estimates of achievable rates in spite of interference.

### 5.7.2 Blockage effects and directionality helps noise-limitedness.

In this section, we show that the blockage effects at mmWave along with directionality of transmissions in the $k-$ring deployment enables noise-limited analysis even if the scheduler does not protect interference on bottleneck links. We now assume that links across orthogonal streets have negligible interference, since the path loss exponent can be as large as 10 for the NLOS segments of such links [114,134]. We simulate the performance of the deployment in Fig. 5.5(a) using NNHR and a greedy variant of the popular backpressure scheduler with congestion control on the first hop as in [143]. We call this as the greedy PF scheduler and it greatly simplifies the implementation of the GBD algorithm in Section I-C of [143]. We choose this particular baseline algorithm, since it emulates PF for multi-hop networks with the utility function in Section I-C of [143] being $U(x) = \log(x)$, which has been a popular paradigm for employing in 4G cellular networks. Another reason for choosing this scheduler is that the discussion in this work is limited to full buffer assumption until now, and considering a scheduler that works for time varying traffic is desirable. This would pave a way for evaluating packet latencies in multi-hop mmWave networks. However, in this section we assume the fiber site always has infinite backlogged data for all UEs. Each BS in the k-ring deployment now represents a queue with multiple traffic flows, each UE repre-




(a) Per UE Access SINR/SNR. (b) Per UE Backhaul SINR/SNR.

Figure 5.8: SINR vs SNR considering the greedy PF scheduler.

senting a flow. Here, we simulate the queueing network for a reasonably long time to understand whether directionality and blockage effects helps keep the network noise-limited even with the proposed simplified scheduler which is not interference-aware. Understanding the stability of the queueing network is an avenue of future research.

Assuming NNHR, scheduling is done as follows. We assign priority of scheduling *a particular flow on a particular link* by the *backpressure* metric, that is product of noise-limited estimate of the instantaneous rate on the link times difference in queue length at the source and destination for the flow on that link. For full buffer traffic, the queue length at source is infinite, which is why the backpressure metric needs modification on the first hop. Congestion control is done as discussed in [143] to emulate PF scheduling and the priority metric for flow $f$ on link $i$, which corresponds to first hop for flow

$f$, is $r_i(t)\left(\frac{1}{cR_i^f(t)} - q_i^f(t)\right)$, where $r_i(t)$ is the noise-limited instantaneous rate of link $i$ at time $t$, $q_i^f$ is the queue length for flow $f$ at the destination of link $i$ and $c$ is the congestion control parameter (set to be $10^{-14}$ to create high backpressure on the first hop for all UEs). Here, $R_i^f(t) = \beta R_i^f(t-1) + (1-\beta)\delta_i^f(t-1)r_{i,\text{actual}}^f(t-1)$, with $\beta = 0.99$ and $\delta_i^f(t-1)$ being the indicator that link $i$ was scheduled for flow $f$ in time slot $t-1$. Here, $r_{i,\text{actual}}^f(t-1)$ is the actual data rate of the scheduled flow $f$ on link $i$ at time $t-1$ (considering interference that resulted as an outcome of the scheduling decision in the previous time slot). Under the assumption of NNHR, scheduling is done using the computed priorities as follows. We pack in links with at least one non-zero priority flow in descending order of the highest priority flow through a link, respecting the half duplex constraint of the devices. If a link is scheduled as per this criterion, then the flows corresponding to highest priority on the scheduled links are chosen in that particular slot for scheduling. This is a greedy variant of the algorithm considered in [143] since instead of searching over all possible transmission schedules we pick a greedy schedule in descending order of priorities.

Fig. 5.7(b) shows the distribution of achieved per user e2e rates for the topology in Fig. 5.5(a), which is computed by dividing the total number of bits received by the UEs from the MBS during a simulation run of 10000 iterations with 0.2 ms slot duration. Data for SINR/SNR and e2e rate was collected after 1000 warm-up iterations and queues at the relays were empty initially. Two blockage scenarios were considered. Scenario 1 implies all links along the same street are LOS to create a worst case interference scenario. Scenario 2

implies only neighbouring backhaul links are LOS, and rest are NLOS. This is reasonable since the non-neighbouring BSs are at least 400m apart, which will likely lead to NLOS links [24, 114]. Scenario 2 also assumes that access links are all NLOS to generate a scenario with low access SNRs. Fig. 5.8 shows the access and backhaul SINR versus SNR comparison in the two blockage scenarios. It can be seen from Fig. 5.7(b) that the rate distribution is almost vertical, implying equal rate per UE was achieved. Ignoring interference, the rate is exactly equal to the max-min rate from our analysis (also equal to 137 Mbps as in Table 5.2), which is surprising at first but can be explained as follows. The bottleneck links for all UEs are those connected to the fiber site having a constant rate $R_1$. Thus, irrespective of whether we do PF or max-min fair scheduling the rates coincide[11]. Note that there is a small drop in rates (by 10%) with interference when all interferers on the same street are LOS. In blockage scenario 2, it is found that the rates with interference do not change at all. This confirms the noise-limited behaviour of the network under consideration.

An intuitive observation that explains noise-limitedness for DL scenario is as follows. All backhaul links operate in a direction away from the fiber site. This along with NNHR ensures that the bottleneck backhaul links connected to the fiber site never see interference with beams aligned towards their receivers from other backhaul links. Nearby access links pointing towards the ring 1 are

[11]Since UE rates on the fiber backhaul links add up to a constant, the solution of PF is same as max-min.

also rarely activated since the ring 2 relays have to backhaul traffic for ring 3 relays. This leads to low interference on bottleneck links.

We note that a maximum limit on SE and SNR inherently provides interference protection for very high values of SINR, even if the actual interference is not negligible. With dense deployment and large antenna arrays (say, $D \leq 200$m, $N_{BS} \geq 64$ and $N_{UE} \geq 16$), this phenomenon may occur routinely, especially in LOS environments. However, we have checked that even if $SE_{max}$ and $SNR_{max}$ are set to be impractically high, the insights in Section 5.7 do not change. An exception would be the noise-limitedness insight for blockage scenario 1 may not hold. This makes accurate modeling of the blockages and directionality crucial to get back noise-limitedness when schedulers do not protect the bottleneck links as in Table 5.2.

## 5.8 Summary

In this chapter, a simple method to compute max-min rates in self-backhauled mmmWave networks was proposed. Apart from the simplicity in the derived formulae and their utility to provide design insights, a key takeaway is that noise-limitedness in the $k-$ring deployment model is aided by the observation that there are a few bottleneck links in the network making it sufficient for an optimal max-min scheduler to activate only a few links at a time. Utility of the analysis as debugging/benchmarking tool for complex system simulators focusing on proportional fairness is also shown through an example.

## 5.9 Appendices

### 5.9.1 Proof of Theorem 9.

Let $\gamma$ be the minimum e2e rate that all UEs can achieve. Knowing the instantaneous rates of the links and the loads, let us now write down necessary conditions for $\gamma$ to be minimum achievable e2e rate, assuming NNHR. The following inequality needs to hold considering the scheduling done by the MBS.

$$\gamma \left( \frac{w_{0,0}}{R_a} + \frac{f(0,0) - w_{0,0}}{R_1} \right) \leq 1. \tag{5.1}$$

Here, left hand side is the total time a BS is active (either transmitting or receiving) and right hand side is the total available time. Here, $\frac{\gamma}{R_a}$ is the minimum fraction of time utilized by MBS for serving a UE directly connected to it on access link. Since there are $w_{0,0}$ such UEs, $\frac{\gamma w_{0,0}}{R_a}$ is the minimum fraction of time MBS spends on access links. Similarly, $\frac{\gamma}{R_1}$ is the minimum fraction of time the MBS spends to serve any indirectly connected user by wireless backhauling. Since there are $f(0,0) - w_{0,0}$ such users, we have the required inequality since fraction of time MBS is active is less than or equal to 1.

Similarly, one can write down inequalities considering minimum fraction of time other BSs need to be active to allow $\gamma$ as the minimum achievable rate to all UEs. Considering the BS at $(i,j)$, with at least $i$ or $j$ not equal to 0, the following inequality can be written.

$$\gamma \left( \frac{w_{i,j}}{R_a} + \frac{f(i,j) - w_{i,j}}{R_1} + \frac{f(i,j)}{R_1} \right) \leq 1,$$

where $\frac{f(i,j)-w_{i,j}}{R_1}$ is the minimum fraction of time the BS has to allocate for backhauling to relays connected to it, further away from $(0,0)$, and $\frac{f(i,j)}{R_1}$ is

the minimum fraction of time the BS is served by its parent BS towards the MBS. Since $w_{i,j} = w_{-i,-j}$ and NNHR, we have $f(i,j) = f(-i,-j)$. Thus, the inequality can be written down as

$$\gamma \left( \frac{w_{i,j}}{R_a} + \frac{f(-i,-j) + f(i,j) - w_{i,j}}{R_1} \right) \leq 1. \tag{5.2}$$

Since $w_{0,0} > w_{i,j}$ and $f(i,j)+f(-i,-j)-w_{i,j} \leq f(i,j)+f(-i,-j) \leq f(0,0)-w_{0,0}$, the inequality (5.1) is stricter than (5.2). Thus, the bottleneck inequality is always (5.1) and thus, $\gamma \leq \left( \frac{w_{0,0}}{R_a} + \frac{f(0,0)-w_{0,0}}{R_1} \right)^{-1}$.

If we prove that a scheduler with NNHR helps achieve the above upper bound, then $\gamma^*$ is the max-min rate. Consider the scheduler in Algorithm 1. If a UE is connected to the MBS, it is clear from the algorithm that its long term rate is $\gamma^*$ since the user gets $\gamma^*/R_a$ fraction of time with instantaneous rate $R_a$. If a UE is connected to the BS at $(i', j')$, then to ensure its long term rate is $\gamma^*$ we need all the backhaul hops to support at least $\gamma^*$ long term rate for the data of this particular user. Also we need the long term access rate for the UE to be at least $\gamma^*$. Since Algorithm 1 allocates $\gamma^* f(i,j)/R_1$ fraction of total time for serving a backhaul link with destination $(i,j)$ and this time is equally divided amongst $f(i,j)$ UEs, the long term rate for any UE amongst the $f(i,j)$ UEs served on this link equals $\gamma^*$. Similarly, the fraction of time any user gets for access is at least $\gamma^*/R_a$, which implies long term rate of $\gamma^*$. Thus, the upper bound $\gamma^*$ is achievable and the max-min rate is given by $\gamma^*$ if the routing is NNHR.

Consider any other routing strategy wherein the fiber site activates only

nearest neighbour backhaul links. Note that inequality in (5.1) still needs to be satisfied as $f(0,0)$ is independent of the routing. Thus, if the nearest neighbour highway routes are changed such that the new links added to the routes do not directly connect with the MBS, it does not change the max-min rates. The only way $\gamma^*$ is not global optimal is if a non-nearest neighbour backhaul links is activated by the fiber site and it outperforms NNHR. If possible, let the MBS serve some of the traffic on links that are not just limited to ring 1 relays. The new equal rate to all UEs, $\tilde{\gamma}$, has to satisfy the following inequality. $\tilde{\gamma}\left(\frac{w_{0,0}}{R_a}+\frac{\beta_1(f(0,0)-w_{0,0})}{R_1}+\frac{\beta_2(f(0,0)-w_{0,0})}{R_2}+\ldots+\frac{\beta_k(f(0,0)-w_{0,0})}{R_k}\right) \leq 1$, where $\sum_{q=1}^{k}\beta_q = 1$ and $\beta_1 < 1$. Since $R_1 > R_2 > \ldots > R_k$, we have that $\tilde{\gamma}$ will always be less than that obtained with NNR. Similarly, modifying any of the inequalities in (5.2) to serve some traffic on links with rates $< R_1$ leads to smaller max-min rates as compared to $\gamma^*$.

# Chapter 6

# Conclusions

This dissertation has proposed new modeling methodologies for studying choice of MIMO techniques and resource allocation in self-backhauled mmWave cellular networks. In particular, instead of considering a link level comparison of MIMO techniques, a system level model was proposed using stochastic geometry for studying the performance of multiuser (MU) MIMO. This model was then used to compare with single user beamforming techniques. Another particularly important outcome of the thesis is to show theoretical feasibility of designing multi-hop self-backhauled networks that meet target 5G data rates. Please refer to Section 1.2 for a detailed summary of the key contributions of this thesis.

In this chapter, key high level takeaways from this dissertation are summarized. Then future research directions motivated from the contributions in this dissertation are discussed.

## 6.1 High-level Takeaways

Following are key high-level takeaways on system design based on the theoretical and empirical studies in this thesis.

- Networks can be noise/interference limited depending on bandwidth, beamwidths, and blockage scenarios.
- Effective antenna gains on LOS and NLOS links can be very different.
- MU-MIMO generally outperforms SU MIMO and SU BF with perfect channel state information but care needed with densification as SU BF might be better.
- Dynamic time division duplexing (TDD) is desirable. There are significant gains in mean data rates for some scenarios but insignificant losses in other scenarios.
- Integrated access-backhaul (IAB) offer better data rates than orthogonal access-backhaul (OAB) schemes in general. But gains are limited in scenarios when number of relays per fiber site is low or when OAB is implemented such that backhaul rate per relay is optimized proportional to the load on the relay.
- In urban canyon scenarios where streets form a square lattice, multi-hop backhauling with as low as 1 fiber site per 40-50 relays can provide 100 Mbps per user rates with 2-4 full buffer users per base stations.
- Noise-limitedness in multi-hop mmWave networks can also arise due to small number of bottleneck links in the network, which makes it sufficient for an optimal scheduler to activate only a few links at a time to still be able to meet max-min end to end user rates.

An obvious question is how should a system design engineer use these insights. These insights are derived/observed based on simplified analytical models and real world is ofcourse much more complex. Thus, an engineer is supposed to keep these insights in mind as rules of thumb to potentially simplify their system simulations or prototyping exercises. Given that the industry simulations are very complex, for example see [5] for 5G new radio channel model, having some rules of thumb based on analysis as done in this dissertation is of potential use to the system design engineers to drive their research with more mathematical understanding and intuition.

## 6.2 Future Research Directions.

### 6.2.1 End to End Rate Performance with Advanced MIMO Techniques.

In Chapter 5, the performance of multi-hop self-backhauled networks was studied considering single sector access points employing single user beamforming. A natural extension would be to study the end to end rate performance considering advanced MIMO techniques implemented by the relays and the fiber sites, possibly having multiple sectors per access point as envisioned for 5G deployments [101]. This will help understanding how large can the value of $k$, in the proposed $k-$ring deployment, be pushed to still support 100 Mbps per user. For example, instead of $k = 3$ rings being the maximum number of rings supported for a given user load, it may be possible that by employing MU-MIMO on backhaul links the limit on $k$ increases to 4 or 5. It would be

interesting to understand whether our insight on applicability of noise-limited analysis from Chapter 5 holds or not. The optimization framework in [43] can be trivially extended for MU-MIMO employed only through analog beam-steering (that is no multiuser interference cancellation, which can be justified since it was observed in Chapter 2 that the zero forcing penalty is close to 1) by modifying the definition feasible schedules. This can be used to compare against a new noise-limited analytical framework on $k-$ring deployment assuming MU-MIMO or multiple sectors per access point.

Since backhaul links are stationary, that is there is no mobility of both the transmitter and the receiver, it can be possible to exploit spatial multiplexing gains even in LOS conditions using LOS MIMO techniques [144]. A comparison of such LOS MIMO techniques versus MU-MIMO performed by the base stations will be interesting to consider. This can be done considering a $k-$ring deployment model as in Chapter 5 and assuming noise-limiteness first. The validity of noise-limitedness can then be studied as done in Chapter 5, considering the impact of large bandwidth, narrow beams, blockages as well as intelligent schedulers which protect interference on bottleneck links.

### 6.2.2 Robustness to Backhaul Links Failures

In Chapter 5, all backhaul hops were assumed LOS. A natural question is that how does the end to end performance change if one or more of the backhaul links are blocked temporarily or permanently. Considering that redeploying the relays is not a feasible solution, there can be two main ways

to tackle this. One is to live with the NLOS backhaul link and re-align the beams to get as high link rate as possible, possibly using advanced MIMO techniques like spatial multiplexing. Another way is to re-route the traffic over that particular link on a different path. The choice of tackling the link failures may also be dependent on whether the link is blocked permanently or temporarily, considering that both schemes will likely require different amount of control signaling that needs to be done. A desirable extension of our work in Chapter 5 is to develop a mathematical model to understand the pros and cons of the above two mechanisms for tackling backhaul link failures. There has been some work on identifying how to self-organize the backhaul network under link failures [145,146]. However, either complete link failures is assumed, that is zero rate on NLOS links, or the solution is in form of a integer linear programming formulation which may be hard to implement in practise. Thus, it will be desirable to see if banking on noise-limited, as in Chapter 5, if there exists a simple solution that can be analytically quantified as well.

### 6.2.3 Robustness to Access Link Failures

Due to dynamic blocking, it is possible for the access links to toggle between LOS and NLOS during the data transmission phase as shown in Fig. 6.1. Thus, it may be desirable to have a backup connection for countering the access link failures. There are several ways to do this and it is unclear which is the best way as a function of different network setups. One method is to have a backup sub-6GHz with the same BS which provides mmWave access.

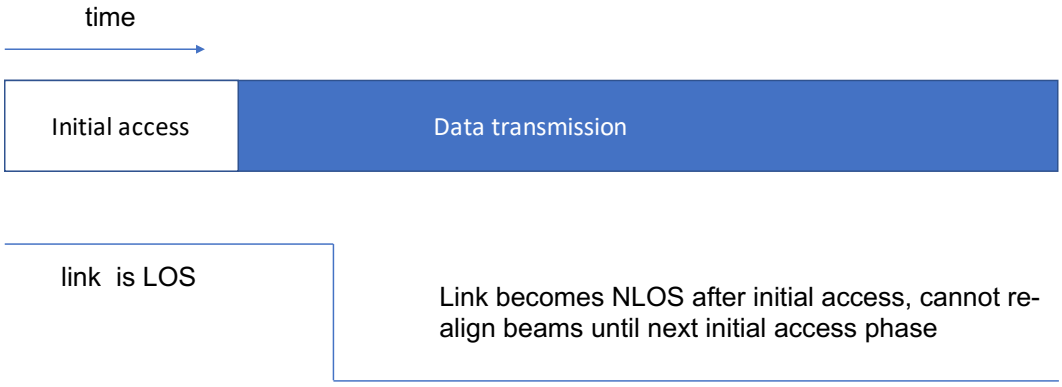



Figure 6.1: Motivation for dual connectivity in mmWave cellular.

Here the assumption is that every BS and UE has a mmWave as well as a sub-6GHz radio. Another method is to have UEs connect to multiple mmWave BSs in the hope that even if one of them is blocked the performance does not suffer. This can itself be done in several ways as discussed in [37]. Developing mathematical models to understand the robustness of mmWave cellular performance to access link failures is another avenue of future research. The choice of robustness mechanism will also be dependent on mobility of user equipments. Although incorporating mobility in rate analysis of cellular networks has been difficult in the past, probably noise-limitedness assumption for mmWave networks may enable some tractable analysis. It would be desirable to model handover penalties in rate as a function of different handover, multi-connectivity mechanisms as well as the velocity of the UEs. In the context of self-driving cars, it may be possible for UEs to share their local travel path with the network making it easier for the network to optimize the multi-connectivity and handover mechanisms.

# Bibliography

[1] M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "A comparison of MIMO techniques in downlink millimeter wave cellular networks with hybrid beamforming," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1952 – 1967, May 2016.

[2] M. N. Kulkarni, E. Visotsky, and J. G. Andrews, "Correction factor for analysis of MIMO wireless networks with highly directional beamforming," *IEEE Wireless Commun. Lett.*, 2018, to appear. Available at https://arxiv.org/abs/1710.07369.

[3] M. N. Kulkarni, J. G. Andrews, and A. Ghosh, "Performance of dynamic and static TDD in self-backhauled millimeter wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6460–6478, Oct 2017.

[4] M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "How many hops can self-backhauled millimeter wave cellular networks support?" *IEEE Trans. Wireless Commun.*, 2018, submitted.

[5] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," TR 38.901, Sept. 2017, version 14.2.0.

[6] M. Weiser, "The computer for the 21st century," *Scientific American*, vol. 265, no. 3, pp. 94–104, September 1991. [Online]. Available: http://doi.acm.org/10.1145/329124.329126

[7] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2787 – 2805, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128610001568

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

[9] "Making 5G NR a reality," Qualcomm, Tech. Rep., Dec. 2016. [Online]. Available: https://bit.ly/2Fyahwi

[10] T. H. J. Collete and B. A. MacDonald, "An augmented reality debugging system for mobile robot software engineers," *Journal of Software Engineering for Robotics*, vol. 1, no. 1, pp. 18–32, Jan. 2010.

[11] D. Kaminska, T. Sapinski, N. Aitken, A. D. Rocca, M. Baranska, and R. Wietsma, "Virtual reality as a tool in mechatronics education," in *Proc. International Symposium on Electromagnetic Fields in Mechatronics, Electrical and Electronic Engineering (ISEF)*, Sept 2017, pp. 1–2.

[12] T. H. J. Collett and B. A. MacDonald, "Augmented reality visualisation for player," in *Proc. IEEE International Conference on Robotics and Automation*, May 2006, pp. 3954–3959.

[13] H. Chen, R. Abbas, P. Cheng, M. Shirvanimoghaddam, W. Hardjawana, W. Bao, Y. Li, and B. Vucetic, "Ultra-reliable low latency cellular networks: Use cases, challenges and approaches," *ArXiv e-prints*, Sept. 2017, available at arXiv:1709.00560.

[14] M. Bennis, M. Debbah, and H. V. Poor, "Ultra-reliable and low-latency wireless communication: Tail, risk and scale," *ArXiv e-prints*, Jan. 2018, available at arXiv:1801.01270.

[15] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, Feb 2014.

[16] "The 5G mmWave revolution," Nokia, Tech. Rep., 2016, available at https://goo.gl/6XBxks.

[17] W. Roh, J. Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

[18] R. G. Fellers, "Millimeter waves and their applications," *Electrical Engineering*, vol. 75, no. 10, pp. 914–917, Oct 1956.

[19] B. Walke and R. Briechle, "A local cellular radio network for digital voice and data transmission at 60 GHz," *Proceedings Cellular & Mobile Comms. Intern.*, pp. 215–225, Nov. 1985.

[20] Y. Li, J. G. Andrews, F. Baccelli, T. D. Novlan, and C. J. Zhang, "Design and analysis of initial access in millimeter wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6409–6425, Oct 2017.

[21] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, March 2014.

[22] V. Va, H. Vikalo, and R. W. Heath, "Beam tracking for mobile millimeter wave communication systems," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec 2016, pp. 743–747.

[23] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, April 2016.

[24] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, June 2014.

[25] S. Hur, T. Kim, D. Love, J. Krogmeier, T. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 4391–4403, Oct. 2013.

[26] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[27] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Jan. 2014.

[28] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace mimo for millimeter-wave communications: System architecture, modeling, analysis and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, July 2013.

[29] J. Singh, S. Ponnuru, and U. Madhow, "Multi-gigabit communication: the ADC bottleneck," in *Proc. IEEE International Conference on Ultra-Wideband*, Sept 2009, pp. 22–27.

[30] A. Alkhateeb, J. Mo, and R. W. Heath, "MIMO precoding and combining solutions for millimeter wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

[31] X. Zhang, A. F. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, Nov 2005.

[32] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[33] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[34] T. Bai and R. W. Heath Jr., "Coverage and rate analysis for millimeter wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 1100–1114, Oct. 2014.

[35] M. N. Kulkarni, A. O. Kaya, D. Calin, and J. G. Andrews, "Impact of humans on the design and performance of millimeter wave cellular networks in stadiums," in *IEEE Wireless Communications and Networking Conference (WCNC)*, March 2017, pp. 1–6.

[36] A. O. Kaya, D. Calin, and H. Viswanathan, "Ultra-broadband, hybrid high-low band wireless access," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, March 2017, pp. 1–6.

[37] M. Polese, M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "Improved handover through dual connectivity in 5G mmWave mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2069–2084, Sept 2017.

[38] Q. C. Li, H. Niu, G. Wu, and R. Q. Hu, "Anchor-booster based heterogeneous networks with mmWave capable booster cells," in *IEEE Globecom Workshops (GC Wkshps)*, Dec 2013, pp. 93–98.

[39] S. Jin, J. Liu, X. Leng, and G. Shen, "Self-backhaul method and apparatus in wireless communication networks," U.S. Patent US20 070 110 005 A1, 2007.

[40] J. García-Rois, F. Gomez-Cuba, M. R. Akdeniz, F. J. Gonzalez-Castano, J. C. Burguillo, S. Rangan, and B. Lorenzo, "On the analysis of scheduling in dynamic duplex multihop mmWave cellular systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6028–6042, Nov. 2015.

[41] M. N. Islam, S. Subramanian, and A. Sampath, "Integrated access backhaul in millimeter wave networks," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, 2017.

[42] J. Du, E. Onaran, D. Chizhik, S. Venkatesan, and R. A. Valenzuela, "Gbps user rates using mmWave relayed backhaul with high gain antennas," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1363 – 1372, June 2017.

[43] M. E. Rasekh, D. Guo, and U. Madhow, "Interference-aware routing and spectrum allocation for millimeter wave backhaul in urban picocells," in *Proc. Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sept 2015, pp. 1–7.

[44] Y. Li, J. Luo, W. Xu, N. Vucic, E. Pateromichelakis, and G. Caire, "A joint scheduling and resource allocation scheme for millimeter wave heterogeneous networks," in *Proc. IEEE WCNC*, March 2017, pp. 1–6.

[45] D. Yuan, H.-Y. Lin, J. Widmer, and M. Hollick, "Optimal joint routing and scheduling in millimeter-wave cellular networks," *Proc. IEEE INFOCOM*, 2018, to appear.

[46] S. Singh, M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "Tractable model for rate in self-backhauled millimeter wave cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2196–2211, Oct. 2015.

[47] J. G. Andrews, T.Bai, M. N. Kulkarni, A. Alkhateeb, A. Gupta, and R. W. Heath, "Modeling and analyzing millimeter wave cellular systems," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 403 – 430, Jan. 2017.

[48] H. Shokri-Ghadikolaei, C. Fischione, G. Fodor, P. Popovski, and M. Zorzi, "Millimeter wave cellular networks: A MAC layer perspective," *IEEE Trans. Commun.*, vol. 63, no. 10, pp. 3437–3458, Oct 2015.

[49] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sept. 2015.

[50] M. Zhang, M. Mezzavilla, J. Zhu, S. Rangan, and S. Panwar, "TCP dynamics over mmWave links," in *Proc. International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, July 2017, pp. 1–6.

[51] J. Gettys and K. Nichols, "Bufferbloat: Dark buffers in the internet," *Queue*, vol. 9, no. 11, pp. 40:40–40:54, Nov. 2011. [Online]. Available: http://doi.acm.org/10.1145/2063166.2071893

[52] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

[53] M. Di Renzo, "Stochastic geometry modeling and analysis of multi-tier millimeter wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 5038–5057, Sept. 2015.

[54] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, 2002.

[55] S. Sun, T. S. Rappaport, R. W. Heath, A. Nix, and S. Rangan, "MIMO for millimeter wave wireless communications: beamforming, spatial mul-

tiplexing, or both?" *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 110–121, Dec. 2014.

[56] A. Ghosh, T. A. Thomas, M. C. Cudak, R. Ratasuk, P. Moorut, F. W. Vook, T. S. Rappaport, G. R. MacCartney, S. Sun, and S. Nie, "Millimeter wave enhanced local area systems: A high data rate approach for future wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1152–1163, June 2014.

[57] J. Mo and R. W. Heath, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5498–5512, Oct. 2015.

[58] A. Sayeed and N. Behdad, "Continuous aperture phased MIMO: Basic theory and applications," *in Proc. Allerton Conference on Communications, Control and Computers*, pp. 1196–1203, Sep. 2010.

[59] F. W. Vook, T. Thomas, and E. Visotsky, "Massive MIMO for mmWave systems," *in Proc. Asilomar Conference on Signals, Systems and Computers*, Nov. 2014.

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

[61] H. S. Dhillon, M. Kountouris, and J. G. Andrews, "Downlink MIMO HetNets: Modeling, ordering results and performance analysis," *IEEE*

*Trans. Wireless Commun.*, vol. 12, no. 10, pp. 5208–5222, Sept. 2013.

[62] H. Zhuang and T. Ohtsuki, "A model based on Poisson point process for analyzing MIMO cellular networks utilizing fractional frequency reuse," *IEEE Trans. Wireless Commun.*, vol. 13, no. 12, pp. 6839–6850, Dec. 2014.

[63] A. K. Gupta, H. S. Dhillon, S. Vishwanath, and J. G. Andrews, "Downlink multi-antenna heterogeneous cellular network with load balancing," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 1830–1844, Nov. 2014.

[64] M. K. Samimi and T. S. Rappaport, "Ultra-wideband statistical channel model for non line of sight millimeter wave urban channels," *IEEE Globecom*, pp. 3483–3489, Dec. 2014.

[65] ——, "Local multipath model parameters for generating 5G millimeter-wave 3GPP-like channel impulse response," *submitted to the 10th European Conference on Antennas and Propagation*, April 2016, Available on ArXiv:1511.06941.

[66] T. Bai and R. W. Heath, "Asymptotic SINR for millimeter wave massive MIMO cellular networks," *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 620–624, June 2015.

[67] M. N. Kulkarni, S. Singh, and J. G. Andrews, "Coverage and rate trends

in dense urban millimeter wave cellular networks," *Proc. IEEE Globecom*, Dec. 2014.

[68] O. E. Ayach, R. W. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, "The capacity optimality of beam steering in large millimeter wave MIMO systems," *IEEE SPACWC*, pp. 100–104, June 2012.

[69] R. Mendez-Rial, C. Rusu, A. Alkhateeb, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "Channel estimation and hybrid combining for mmWave: Phase shifters or switches?" *Proc. of Information Theory and Applications Workshop*, Feb. 2015.

[70] A. Adhikary, E. A. Safadi, M. Samimi, R. Wang, G. Caire, T. S. Rappaport, and A. F. Molisch, "Joint spatial division and multiplexing for mmWave channels," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1239–1255, May 2014.

[71] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, November 2010.

[72] F. Baccelli and B. Blaszczyszyn, *Stochastic Geometry and Wireless Networks, Volume I – Theory.* NOW: Foundations and Trends in Networking, 2009.

[73] S. Singh, H. S. Dhillon, and J. G. Andrews, "Offloading in heterogeneous networks: Modeling, analysis, and design insights," *IEEE Trans.*

*Wireless Commun.*, vol. 12, no. 5, pp. 2484–2497, May 2013.

[74] J.-S. Ferenc and Z. Néda, "On the size distribution of Poisson Voronoi cells," *Physica A: Statistical Mechanics and its Applications*, vol. 385, no. 2, pp. 518 – 526, Nov. 2007.

[75] B. Blaszczyszyn, M. K. Karray, and H.-P. Keeler, "Using Poisson processes to model lattice cellular networks," in *Proc. IEEE INFOCOM*, Apr. 2013, pp. 773–781.

[76] C. Balanis, *Antenna Theory*. Wiley, 1997.

[77] T. Bai, A. Alkhateeb, and R. W. Heath, "Coverage and capacity of millimeter-wave cellular networks," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 70–77, Sept. 2014.

[78] M. N. Kulkarni, A. Alkhateeb, and J. G. Andrews, "A tractable model for per user rate in multiuser millimeter wave cellular networks," *in Proc. IEEE Asilomar Conference on Signals, Systems and Computers*, Nov. 2015.

[79] J. Yu, Y. D. Yao, J. Zhang, and A. F. Molisch, "Reverse link capacity of power-controlled CDMA systems with antenna arrays in a multipath fading environment," *IEEE Globecom*, pp. 839–843, Dec. 2003.

[80] A. M. Hunter, J. G. Andrews, and S. Weber, "Transmission capacity of ad hoc networks with spatial diversity," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5058–5071, Dec. 2008.

[81] J. Wildman, P. H. J. Nardelli, M. Latva-aho, and S. Weber, "On the joint impact of beamwidth and orientation error on throughput in directional wireless Poisson networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 12, pp. 7072–7085, Dec 2014.

[82] D. Maamari, N. Devroye, and D. Tuninetti, "Coverage in mmWave cellular networks with base station co-operation," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2981–2994, April 2016.

[83] J. Park, S. L. Kim, and J. Zander, "Tractable resource management with uplink decoupled millimeter-wave overlay in ultra-dense cellular networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4362–4379, June 2016.

[84] N. Deng and M. Haenggi, "A fine-grained analysis of millimeter-wave device-to-device networks," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4940–4954, Nov 2017.

[85] D. Chizhik, G. J. Foschini, and R. A. Valenzuela, "Capacities of multi-element transmit and receive antennas: Correlations and keyholes," *Electronics Letters*, vol. 36, no. 13, pp. 1099–1100, Jun 2000.

[86] M. Lugo, "The expectation of the maximum of exponentials," *Stat 134 Class Notes, UC Berkeley*, Oct. 2011, available at https://goo.gl/RGVxiR.

[87] R. Taori and A. Sridharan, "Point-to-multipoint in-band mmwave backhaul for 5G networks," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 195–

201, January 2015.

[88] J. Li, S. Farahvash, M. Kavehrad, and R. Valenzuela, "Dynamic TDD and fixed cellular networks," *IEEE Commun. Lett.*, vol. 4, no. 7, pp. 218–220, July 2000.

[89] Z. Shen, A. Khoryaev, E. Eriksson, and X. Pan, "Dynamic uplink-downlink configuration and interference management in TD-LTE," *IEEE Commun. Mag.*, vol. 50, no. 11, pp. 51–59, Nov. 2012.

[90] A. K. Gupta, M. N. Kulkarni, E. Visotsky, F. W. Vook, A. Ghosh, J. G. Andrews, and R. W. Heath, "Rate analysis and feasibility of dynamic TDD in 5G cellular systems," *in Proc. IEEE ICC*, pp. 1–6, 2016.

[91] Nokia. (2015) 5G radio access system design aspects. Available at: https://goo.gl/99POAV.

[92] Ericsson. (2016) 5G radio access- capabilities and technologies. Available at https://goo.gl/3vsqVy.

[93] B. Yu, L. Yang, H. Ishii, and S. Mukherjee, "Dynamic TDD support in macrocell-assisted small cell architecture," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1201–1213, June 2015.

[94] H. Sun, M. Wildemeersch, M. Sheng, and T. Q. S. Quek, "D2D enhanced heterogeneous cellular networks with dynamic TDD," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4204–4218, Aug 2015.

[95] E. Dahlman, S. Parkvall, and J. Sköld, *4G LTE/LTE-Advanced for Mobile Broadband*, 2nd ed. Academic Press, 2014.

[96] H. Viswanathan and S. Mukherjee, "Performance of cellular networks with relays and centralized scheduling," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2318–2328, Sept 2005.

[97] A. Ghosh (AT&T), "Designing ultra-dense networks for 5G," in *Texas Wireless Summit*, 2016. [Online]. Available: https://youtu.be/Xp2fGEzD5JQ

[98] W. Lu and M. D. Renzo, "Stochastic geometry modeling and system-level analysis & optimization of relay-aided downlink cellular networks," *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 4063–4085, Nov 2015.

[99] H. Tabassum, A. H. Sakr, and E. Hossain, "Analysis of massive MIMO-enabled downlink wireless backhauling for full-duplex small cells," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2354–2369, June 2016.

[100] A. Sharma, R. K. Ganti, and J. K. Milleth, "Joint backhaul-access analysis of full duplex self-backhauling heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1727–1740, March 2017.

[101] A. Ghosh, "The 5G mmWave radio revolution," *Microwave Journal*, vol. 59, no. 9, pp. 22–36, Sep. 2016.

[102] H. ElSawy, A. Sultan-Salem, M. S. Alouini, and M. Z. Win, "Modeling and analysis of cellular networks using stochastic geometry: A tutorial,"

*IEEE Communication Surveys and Turorials*, vol. 19, no. 1, pp. 167–203, 2017.

[103] S. M. Yu and S.-L. Kim, "Downlink capacity and base station density in cellular networks," in *Workshop in Spatial Stochastic Models for Wireless Networks*, May 2013.

[104] M. D. Renzo, W. Lu, and P. Guan, "The intensity matching approach: A tractable stochastic geometry approximation to system-level analysis of cellular networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 5963–5983, Sept 2016.

[105] V. Suryaprakash, J. Møller, and G. Fettweis, "On the modeling and analysis of heterogeneous radio access networks using a Poisson cluster process," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 1035–1047, Feb 2015.

[106] Y. Li, F. Baccelli, H. S. Dhillon, and J. G. Andrews, "Statistical modeling and probabilistic analysis of cellular networks with determinantal point processes," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3405–3422, Sept 2015.

[107] S. Singh, X. Zhang, and J. G. Andrews, "Joint rate and SINR coverage analysis for decoupled uplink-downlink biased cell associations in HetNets," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5360–5373, Oct 2015.

[108] H. ElSawy and E. Hossain, "On stochastic geometry modeling of cellular uplink transmission with truncated channel inversion power control," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4454–4469, Aug 2014.

[109] H. Lee, Y. Sang, and K. Kim, "On the uplink SIR distributions in heterogeneous cellular networks," *IEEE Commun. Let.,*, vol. 18, no. 12, pp. 2145–2148, Dec. 2014.

[110] M. Haenggi, "User point processes in cellular networks," *IEEE Wireless Commun. Lett.*, vol. 6, no. 2, April 2017.

[111] M. N. Kulkarni. (2017, Jan.) MATLAB codes for self-backhauled mmWave cellular networks. Available at: https://goo.gl/VoAjWm.

[112] H. Elshaer, M. N. Kulkarni, F. Boccardi, J. G. Andrews, and M. Dohler, "Downlink and uplink cell association with traditional macrocells and millimeter wave small cells," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6244–6258, Sept. 2016.

[113] T. S. Rappaport, R. W. Heath, R. C. Daniels, and J. N. Murdock, *Millimeter wave wireless communications*. Prentice Hall, 2014.

[114] A. Karttunen, A. F. Molisch, S. Hur, J. Park, and C. J. Zhang, "Spatially consistent street-by-street path loss model for 28-GHz channels in micro cell urban environments," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7538–7550, Nov 2017.

[115] B. Hajek and G. Sasaki, "Link scheduling in polynomial time," *IEEE Trans. Inf. Theory*, vol. 34, no. 5, pp. 910–917, Sep 1988.

[116] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Trans. Autom. Control*, vol. 37, no. 12, pp. 1936–1948, Dec 1992.

[117] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *J. Oper. Res. Soc.*, vol. 49, no. 3, pp. 237–252, Mar 1998. [Online]. Available: https://doi.org/10.1057/palgrave.jors.2600523

[118] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," *Wireless Networks*, vol. 11, no. 4, pp. 471–487, Jul 2005.

[119] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, March 2000.

[120] M. Franceschetti, O. Dousse, D. N. C. Tse, and P. Thiran, "Closing the gap in the capacity of wireless networks via percolation theory," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 1009–1018, March 2007.

[121] A. Zemlianov and G. de Veciana, "Capacity of ad hoc wireless networks with infrastructure support," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 3, pp. 657–667, March 2005.

[122] I. Gitman, "On the capacity of slotted ALOHA networks and some design problems," *IEEE Trans. Commun.*, vol. 23, no. 3, pp. 305–317, Mar 1975.

[123] L. Kleinrock and J. Silvester, "Spatial reuse in multihop packet radio networks," *Proc. IEEE*, vol. 75, no. 1, pp. 156–167, Jan 1987.

[124] M. Sikora, J. N. Laneman, M. Haenggi, D. J. Costello, and T. E. Fuja, "Bandwidth- and power-efficient routing in linear wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2624–2633, June 2006.

[125] F. Baccelli, B. Blaszczyszyn, and P. Muhlethaler, "An Aloha protocol for multihop mobile wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 421–436, Feb 2006.

[126] J. G. Andrews, S. Weber, M. Kountouris, and M. Haenggi, "Random access transport capacity," *IEEE Trans. Wireless Commun.*, vol. 9, no. 6, pp. 2101–2111, June 2010.

[127] "Part 11: Wireless LAN MAC and PHY specifications amendment 10: Mesh networking," *IEEE Std 802.11s*, pp. 1–372, Sept 2011.

[128] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Relay radio transmission and reception; Release 11," TR 36.826, 2012.

[129] S. W. Peters and R. W. Heath Jr., "The future of WiMAX: Multihop relaying with IEEE 802.16j," *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 104–111, January 2009.

[130] J. Andrews, S. Shakkottai, R. Heath, N. Jindal, M. Haenggi, R. Berry, D. Guo, M. Neely, S. Weber, S. Jafar, and A. Yener, "Rethinking information theory for mobile ad hoc networks," *IEEE Commun. Mag.*, vol. 46, no. 12, pp. 94–101, Dec. 2008.

[131] R. Mudumbai, S. K. Singh, and U. Madhow, "Medium access control for 60 GHz outdoor mesh networks with highly directional links," in *Proc. IEEE Infocom*, April 2009, pp. 2871–2875.

[132] F. B. Tesema, A. Awada, I. Viering, M. Simsek, and G. P. Fettweis, "Mobility modeling and performance evaluation of multi-connectivity in 5G intra-frequency networks," in *Proc. IEEE Globecom Workshops*, Dec 2015, pp. 1–6.

[133] Z. He, S. Mao, S. Kompella, and A. Swami, "Minimum time length scheduling under blockage and interference in multi-hop mmWave networks," in *Proc. IEEE Globecom*, Dec 2015, pp. 1–7.

[134] Y. Wang, K. Venugopal, A. F. Molisch, and R. W. Heath, "Mmwave vehicle-to-infrastructure communication: Analysis of urban microcellular networks," *ArXiv e-prints*, 2018, available at arXiv:1702.08122.

[135] D. P. Bertsekas and R. Gallagar, *Data Networks*, 2nd ed. Prentice Hall, 1992.

[136] V. Petrov, D. Solomitckii, A. Samuylov, M. A. Lema, M. Gapeyenko, D. Moltchanov, S. Andreev, V. Naumov, K. Samouylov, M. Dohler,

and Y. Koucheryavy, "Dynamic multi-connectivity performance in ultra-dense urban mmWave deployments," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2038–2055, Sept 2017.

[137] Z. Zhang, J. Ryu, S. Subramanian, and A. Sampath, "Coverage and channel characteristics of millimeter wave band using ray tracing," in *Proc. IEEE ICC*, June 2015, pp. 1380–1385.

[138] M. O. Hasna and M. S. Alouini, "Outage probability of multihop transmission over Nakagami fading channels," *IEEE Commun. Lett.*, vol. 7, no. 5, pp. 216–218, May 2003.

[139] Ericsson. (2017) Designing for the future: the 5G NR physical layer. Available at https://goo.gl/jU5S1r.

[140] T. Dinc and H. Krishnaswamy, "Millimeter-wave full-duplex wireless: Applications, antenna interfaces and systems," in *Proc. IEEE Custom Integrated Circuits Conference (CICC)*, April 2017, pp. 1–8.

[141] P. Mogensen, W. Na, I. Z. Kovacs, F. Frederiksen, A. Pokhariyal, K. I. Pedersen, T. Kolding, K. Hugl, and M. Kuusela, "LTE capacity compared to the Shannon bound," in *Proc. IEEE VTC Spring*, April 2007, pp. 1234–1238.

[142] M. N. Kulkarni. (2018, March) MATLAB codes for multi-hop mmWave cellular networks. Available at: https://goo.gl/VoAjWm.

[143] U. Akyol, M. Andrews, P. Gupta, J. Hobby, I. Saniee, and A. Stolyar, "Joint scheduling and congestion control in mobile ad-hoc networks," in *IEEE INFOCOM*, April 2008.

[144] E. Torkildson, U. Madhow, and M. Rodwell, "Indoor millimeter wave MIMO: Feasibility and performance," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4150–4160, December 2011.

[145] Y. H. Chiang and W. Liao, "mw-HierBack: A cost-effective and robust millimeter wave hierarchical backhaul solution for HetNets," *IEEE Trans. Mobile Comput.*, vol. 16, no. 12, pp. 3445–3458, Dec 2017.

[146] G. Kuperman and E. Modiano, "Providing guaranteed protection in multi-hop wireless networks with interference constraints," *IEEE Trans. Mobile Comput.*, vol. 16, no. 12, pp. 3502–3512, Dec 2017.

# Vita

Mandar N. Kulkarni received M.S. in Electrical Engineering from the University of Texas at Austin in 2015 and B.Tech in Electronics and Communications Engineering from the Indian Institute of Technology (IIT) Guwahati in 2013. He has held internship positions at Nokia Bell Labs, Murray Hill, NJ, U.S.A. (2016); Nokia Networks, Arlington Heights, IL, U.S.A.(2014, 2015, 2017); Technical University, Berlin, Germany (2012); and Indian Institute of Science, Bangalore, India (2011). He received the President of India Gold Medal in 2013 for his academic performance at IIT Guwahati, and George J. Heuer Jr. Fellowship at UT Austin for the year 2013-2014.

Permanent email: kulkarni.mandar.n@gmail.com

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.