The Impact of Multi-Connectivity and Handover Constraints on Millimeter Wave and Terahertz Cellular Networks

Mustafa F. Özkoç, *Student Member, IEEE*, Athanasios Koutsaftis, *Student Member, IEEE*, Rajeev Kumar, *Member, IEEE*, Pei Liu, *Member, IEEE*, and Shivendra S. Panwar, *Fellow, IEEE*

***Abstract*— Wireless communication over terahertz (THz) frequency bands is envisioned as the key enabler of many applications and services offered in 6G networks. The abundantly available bandwidth in THz frequencies can satisfy the ultra-high user throughput requirements and accommodate a massive number of connected devices. However, poor propagation characteristics, shadowing, and blockages may result in sudden outages and necessitate frequent handovers. Therefore, an inefficient handover procedure will impose severe challenges in meeting the ultra-high reliability and low latency requirements of emerging applications. In blockage driven mmWave and THz networks, a higher multi-connectivity degree and efficient handover procedures are needed to reduce the data plane interruptions and to achieve high reliability. We present an analytical model to study the impact of handover procedures and multi-connectivity degree on the latency and reliability of blockage driven wireless networks. From the network protocol design perspective, our study offers a quick and accurate way to envisage how network architecture and protocols should evolve in terms of multi-connectivity degrees and handover procedural efficiency. Our results suggest that, for THz systems, coverage range should be increased even if it comes at the cost of increased initial access and base station discovery times.**

***Index Terms*— Handover, reliability, low latency, millimeter wave, mmWave, terahertz, THz, blockages, multi-connectivity, quality of service, URLLC.**

I. Introduction

FIFTH Generation (5G) cellular networks, currently being rolled out, serve a wide range of new emerging applications and services including eHealth, Augmented and Virtual Reality, and the tactile Internet. Not only do these applications and services require a high data rate in the range of 100 Mbps to a few Gbps, but they also impose stringent latency and reliability requirements on the 5G cellular networks. To implement different policies based on the QoS requirements of applications and services, the 3rd Generation Partnership Project (3GPP) categorizes them in three different classes of services [1], [2]. These are massive Machine Type Communication (mMTC), enhanced Mobile Broadband (eMBB), and Ultra-Reliable Low Latency Communication (URLLC). A comprehensive set of requirements for these services and applications is presented in Table I. As throughput is the primary performance metric for eMBB services, 5G wireless networks are expected to use millimeter wave (mmWave) frequencies, due to its large available bandwidth [3], [4].

Although mmWave frequencies are adequate for the current traffic demand in 5G, wireless usage is expected to increase significantly over the next decade [7]. In fact, it is anticipated that by 2030, the wireless data rates will compete with wired broadband [8]. Towards this end, the terahertz (THz) frequency band (0.1-10 THz) is attracting attention from the global community, and its distinctive characteristics are extensively being studied in order to achieve an efficient communication channel for Sixth Generation (6G) networks. Nevertheless, 6G networks are expected to offer services, such as holographic communications, fully automated driving, and telesurgery [9]–[11], as well as applications that have not yet been conceived. These emerging 6G use cases will impose stringent design requirements to jointly satisfy ultra-high throughput, ultra-high reliability, and extremely low latency [12], as distinct from 5G uses cases [1], [2] which do not require high throughput, ultra-high reliability, and low latency simultaneously.

Although mmWave and THz systems are capable of transmitting at speeds of multiple gigabits-per-second on the air interface, they share a vulnerability to blockages and shadowing [13]–[16]. mmWave links can suffer significant penetration loss due to dynamic and static blockers, while the penetration loss at THz frequencies can be up to 3-4 times higher [12]. Furthermore, free-space path loss is higher in THz frequencies, where the attenuation is proportional to the square of carrier frequency [17]. As a result, the combined attenuation in the THz band is considerably high, limiting the maximum transmission range of future devices [18]. To overcome such high path loss, a highly directive antenna pattern must be used, which however results in frequent misalignment of beams due to small scale mobility (UEs) [19]. High directivity is achieved by narrow beam-vectors steered by a large number of antenna

Manuscript received July 7, 2020; revised November 22, 2020 and February 20, 2021; accepted March 1, 2021. Date of publication April 8, 2021; date of current version May 18, 2021. This work was supported by NYU Wireless, an Ernst Weber Fellowship, a gift from Futurewei Technologies and in part by the NY State Center for Advanced Technology in Telecommunications (CATT). *(Corresponding author: Mustafa F. Özkoç.)*

Mustafa F. Özkoç, Athanasios Koutsaftis, Pei Liu, and Shivendra S. Panwar are with the Department of Electrical and Computer Engineering, Tandon School of Engineering, New York University, Brooklyn, NY 11210 USA (e-mail: ozkoc@nyu.edu; tkoutsaftis@nyu.edu; peiliu@nyu.edu; panwar@nyu.edu).

Rajeev Kumar was with the Department of Electrical and Computer Engineering, Tandon School of Engineering, New York University, Brooklyn, NY 11210 USA. He is now with Qualcomm Inc., San Diego, CA 92121 USA (e-mail: rajeevkr@nyu.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSAC.2021.3071852.

Digital Object Identifier 10.1109/JSAC.2021.3071852

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

TABLE I
QoS REQUIREMENTS AND EXAMPLE APPLICATIONS FOR mMTC, URLLC AND eMBB SERVICES IN 5G [1], [5], [6]. UNLIKE 5G, 6G APPLICATIONS WILL LIKELY SIMULTANEOUSLY REQUIRE HIGH RELIABILITY, HIGH BANDWIDTH AND LOW LATENCY

Services/Application Category	Throughput	Air-link Latency	Reliability	Example Applications
mMTC	1-100 Kbps	10 ms - 1 hr	90%	smart city, smart home
URLLC	1-10 Mbps	1 ms	99.9 % - 99.9999 %	eHealth, factory automation, robotics
eMBB	0.1 - 10 Gbps	4 ms	99.9 %	AR, VR, tactile Internet, 360 degrees video

elements, which requires a lengthy period to achieve beam alignment [20], [21]. As a result, the cell search time, or equivalently the Base Station (BS) discovery time, can be significantly higher for THz systems as compared to mmWave systems. Although frequent blockages exist in both mmWave and THz systems, the impact on performance during handover can be significantly worse in THz systems due to the higher BS discovery time. In addition to the higher BS discovery time, the high cell association time due to inefficient handover mechanisms [22], or equivalently the high handover execution time will induce a further increase in handover latency, which will be unacceptably high for 6G networks. Moreover, service reliability can severely deteriorate if handover is not performed in a timely fashion, due to handover failures and Radio Link Failures (RLFs) [22], [23]. Fast discovery of unblocked BSs as well as robust and efficient handover techniques will be required to exploit the full promise of THz cellular networks. This necessitates an extensive study of the impact of BS discovery time and handover latency on the QoS of blockage driven cellular networks.

As mmWave and THz links are less reliable and intermittent, it will be greatly beneficial for the UEs to harness macrodiversity from the nearby BSs in 5G and future cellular networks. Concomitantly, networks should evolve to support fast BS discovery, efficient handover procedures, and efficient mechanisms for achieving multi-connectivity. Therefore, it is essential to study the impact of the *BS discovery time*, the *handover execution time*, and the *degree of multi-connectivity* on the QoS of different applications and services. As discussed earlier, mmWave and THz systems may have different propagation characteristics, yet they are both vulnerable to static and dynamic blockages, thus a similar handover analysis framework can be applicable to both of these blockage driven cellular networks. A comprehensive study of the trade-offs among the aforementioned parameters will be vital for the design of next-generation mmWave and THz cellular networks.

In this paper, we use a stochastic geometry based model and a continuous time Markov chain to study the impact of the BS discovery times (including handover preparation times), handover execution time, and the multi-connectivity degree on the out-of-service probability, the out-of-service duration, and RLF probability. This, in turn, allows us to determine whether and in which ways the network architecture and protocols should further evolve to satisfy the QoS of different applications and services in mmWave and THz cellular networks, where links are highly intermittent in nature. The key contributions of this paper are as follows:

- We provide an analytical model to study the impact of BS discovery time, handover execution time, and the degree of multi-connectivity on the QoS of different applications and services considering dynamic blockages (UE blocked by mobile blockers) and self-blockages (UE blocked by the user's own body).
- We derive closed-form expressions and/or bounds on the out-of-service and RLF probabilities, and expected out-of-service duration using stochastic geometry and Markov chain frameworks. From our analytical study, we derive the required degree of connectivity, BS discovery and handover execution times to meet the QoS requirements of future applications and services for different BS density, coverage range, and blocker density values.
- We verify our analytical results through Monte-Carlo simulations by considering the random waypoint mobility model for the movement of blockers. Finally, using our initial results, we argue that for THz cellular networks, coverage range should be increased to reduce the required BS density even if it comes at the cost of increased BS discovery time. Moreover, a dual connectivity architecture is needed to avoid short interruptions caused by protocol limitations.

The rest of the paper is organized as follows. Section II presents related work. Section III describes the system model. In Section IV, we provide an analytical model to study the impact of handover latency (BS discovery and handover execution times) and multi-connectivity degree on the overall QoS of different applications and services. In Section V, we present results obtained by our theoretical analysis and MATLAB simulations. In Section VI, we discuss the implications of our results on mmWave and THz cellular networks. Finally, Section VII concludes the paper.

II. RELATED WORK

The BS discovery time (measurement reports of 200 ms [24]), handover trigger time (a couple of measurement reports [24]), and handover procedures (Break-Before-Make (BBM) [25]) in legacy Long Term Evolution (LTE) heterogeneous networks are prohibitively high for the 5G and next-generation wireless networks. In the BBM handover mechanism, the UE breaks the existing connection with the serving BS before initiating a handover procedure to the target BS that results in a $40 - 50$ ms interruption to data plane services [22]. In the 3GPP standard (Release 14), *Make-Before-Break (MBB)* and *Random Access Channel (RACH)-less* techniques were introduced to significantly reduce these data plane interruptions during handover [26]. In the MBB handover procedures, the UE breaks the existing connection with the serving BS only before the random access procedure is initiated with the target BS.

Due to the high implementation cost and the limited benefits of the MBB and RACH-less handover techniques in practical

systems [27], Conditional Handover (CHO), and Dual Active Protocol Stack (DAPS) handover techniques were introduced in Release 16 [28] to achieve mobility robustness. In the CHO, the network can configure the UE with multiple target BSs in the RRC Reconfiguration message. Furthermore, the UE can appropriately select a target BS from the configured BSs in the RRC Reconfiguration message when the handover execution criteria are met. If the handover procedures fail with the selected target BS, the UE can select and attempt to connect to another target BS. Although CHO reduces the handover failure probability, it increases the handover latency if the UE performs multiple HO attempts during a single RRC Reconfiguration. In the DAPS handover technique, the connection to the source BS is maintained unless and until the connection to the target BS is fully made. Even though the DAPS handover technique reduces the handover failure probability and downlink data plane interruptions, the uplink data plane is still interrupted during the execution of a DAPS handover [28]. Furthermore, a DAPS handover can still fail if both source and target BSs get blocked.

Another technique that can help in achieving low latency and high reliability in cellular networks is *multi-connectivity*. Dual Connectivity (DC) in heterogeneous LTE networks was introduced in the 3GPP standard Release 12 [26] primarily to achieve high throughput by maximizing the received power at the UE. Although DC contributed to throughput gain, it also resulted in higher buffering latency at the secondary BS [29]. In [30], it was shown that increasing the multi-connectivity degree up to 4 can be beneficial for both outage probability and spectral efficiency. Moreover, it was suggested that any further increase on the multi-connectivity degree does not significantly improve the performance, but instead dramatically increases the signaling overhead and the complexity of the network protocols. However, switching between the BSs was assumed instantaneous in [30], i.e., the BS discovery and handover execution times were not considered in the analysis. In 5G and next-generation cellular networks, where frequent handovers may be required, Multi-Radio Access Technology (Multi-RAT) and multi-connectivity may prioritize low latency and high reliability over high throughput. In our previous work [31], we showed that by prioritizing low latency and high reliability, fast BS discovery, and an efficient handover procedure we can significantly enhance QoS in mmWave cellular networks.

To achieve a high QoS, M. Polese *et al.* [32] have proposed procedures for fast switching among BSs in the multi-RAT DC settings. During a blockage, the UE switches to the LTE BS after receiving a handover command, and once a new mmWave BS is found, the UE switches to the discovered mmWave BS. This results in a temporary overloading of the LTE BSs [33] and a framework for fast switching among mmWave BSs will be required. Considering different multi-connectivity scenarios, V. Petrov *et al.* in [34] studied the impact of the multi-connectivity degree. In our previous work [31], we proposed a new transport network architecture for mmWave cellular systems and fast control signaling among BSs to achieve high QoS for different applications and services. Using MATLAB simulations we showed that the achievable QoS is dependent upon the BS discovery time, handover execution time, and the multi-connectivity degree. In this paper, we want to explicitly establish a relationship between the BS discovery time, the handover execution time, and the multi-connectivity degree with the achieved QoS of different applications and services.

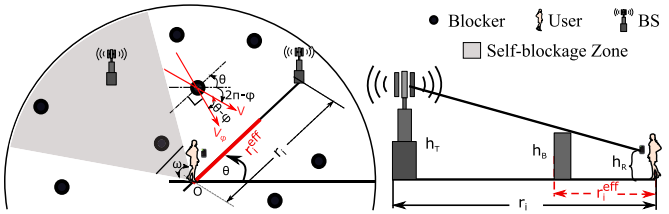
In [35], B. Zhang *et al.* proposed a Markov chain based process to model the handover procedure and offloading in LTE heterogeneous networks. By optimizing different parameters, they attempted to decrease the handover rate and the load imbalance caused by different transmission powers. In an effort to optimize the handover procedure in the LTE heterogeneous networks, F. Guidolin *et al.* [36] proposed a novel context-aware policy. Their scheme aims to maximize the UE average capacity by using a non-homogeneous discrete time Markov chain to model the evolution of the UE state. M. Gerasimenko *et al.* [37] and A. Shafie *et al.* [38] considered Markov chains for developing connectivity strategies in the mmWave and THz network respectively, where they assumed that the UE maintains backup connections with the other non-blocked access points. However, in practical systems, a UE can maintain connections with only a limited number of access points. Furthermore, the UE will always require additional time to find the unblocked access points and associate with a discovered access point.

V. Begishev *et al.* considered the session management in multi-band microwave and mmWave multi-connectivity systems [39] and in 3GPP New Radio (NR) multi-connectivity systems [40], using multi-dimensional Markov chains. In the former work [39], switching between mmWave BS and sub-6 GHz BS depending on the available resources and blockage events is considered. However, the authors assumed immediate switching between sub-6 GHz BS and mmWave BS while neglecting the discovery and association delays for the mmWave BS after blockage events end. In the latter work [40], the authors have considered a single NR BS and demonstrated that reserving guard capacity for the services of already accepted UEs can increase the system throughput while overlooking the reliability and latency requirements of the services.

Authors in [35]–[40] aim to maximize the UE throughput without considering the induced latency and service interruptions due to the blockages in the mmWave and THz cellular networks. In mmWave and THz cellular networks, where abundant bandwidth is available but the channels are sporadic in nature, it will be more challenging to meet the low latency and ultra-high reliability requirements of applications and services than delivering high throughput. Therefore, we focus our study on the reliability and latency of the services by taking into account several key factors such as BS discovery time, handover execution time in addition to the multi-connectivity degree, which is essential for an optimized design of the blockage driven mmWave and THz cellular networks.

III. System Model

In our system model, the BS blocking and unblocking processes are similar to the model in [14] which are presented



(a) Top view (b) Side view

Fig. 1. System model [14]. The physical blockage process for BSs is similar to [14]. The handover latency impact is neglected in [14], where instantaneous handover in a macro diversity setting is considered.

in III-A for the sake of convenience. We discuss the handover procedure related details of our system model in III-B. The novelty of our work lies in considering the blockage process together with the handover procedures to evaluate their impact on blockage driven mmWave and THz cellular networks. A brief summary of notations is provided in Table II.

A. Physical Blockage Process

A stationary UE is assumed at the center of a disc $B(o,R)$ having radius R and origin at o. In a mmWave and THz cellular system, the Line-of-Sight (LOS) paths between the UE and the BS can be blocked by the user's own body and/or dynamic blockers such as humans, moving vehicles, and other moving objects around the UE (see Fig. 1). The system model consists of the following components:

1) BS Model: The mmWave and/or THz BS location is modeled as a homogeneous Poisson Point Process (PPP) inside the disc $B(o,R)$ with density λ_T. The number of BSs M in the disc follows a Poisson distribution with parameter $\lambda_T \pi R^2$:

$$P_M(m) = \frac{\left[\lambda_T \pi R^2\right]^m}{m!} e^{-\lambda_T \pi R^2}. \tag{1}$$

Given the number of BSs in the disc $B(o,R)$, the BSs locations follow a uniform probability distribution. Thus, the radial distance R_i $i = 1, 2, \cdots, m$ of BSs from the UE are independent and identically distributed (i.i.d.):

$$f_{R_i|m}(r_i) = \frac{2r_i}{R^2};\ 0 \le r_i \le R,\ i = 1, 2, \cdots, m, \tag{2}$$

and the angular positions $\theta_i,\ i = 1, 2, \cdots, m$ of BSs relative to the positive x-axis are i.i.d. and follow uniform distribution in $[0, 2\pi]$.

2) Self-Blockage Model: The self-blockage zone is defined as the sector of the disc $B(o,R)$ with angle κ in which all of the BSs are considered blocked to the UE due to the user's own body (see Fig. 1). Note that the orientation of the user's body is uniform in $[0, 2\pi]$. Thus, the probability that a randomly chosen BS is blocked due to self-blockage can be derived as

$$P\left(B^{\text{self}}\right) = \frac{\kappa}{2\pi}. \tag{3}$$

Further, the probability that a randomly chosen BS is not blocked by self-blockage is derived as

$$p = 1 - P\left(B^{\text{self}}\right) = 1 - \frac{\kappa}{2\pi}. \tag{4}$$

Thus, the number of BSs in the disc $B(o,R)$ out of the self-blockage zone follows another Poisson distribution with parameter $p\lambda_T \pi R^2$, i.e.,

$$P_M(m) = \frac{\left[p\lambda_T \pi R^2\right]^m}{m!} e^{-p\lambda_T \pi R^2}. \tag{5}$$

Furthermore, the Line-of-sight (LOS) coverage probability, $P\left(C^{\text{LOS}}\right)$, can be obtained by considering that at least one BS should lie outside the self-blockage zone in disc $B(o,R)$, i.e.,

$$P\left(C^{\text{LOS}}\right) = P_M(m \neq 0) = 1 - e^{-p\lambda_T \pi R^2}. \tag{6}$$

3) Dynamic Blockage Model: Dynamic blockages (illustrated in Fig. 1) in the mmWave cellular networks are extensively studied in [14], [15] assuming a homogeneous PPP model with dynamic blocker density λ_B in the disc $B(o,R)$. The blocker arrival rate α_i at the i^{th} BS-UE link is considered Poisson and was derived in [14], [15] as

$$\alpha_i = \Theta r_i,\ i = 1, 2, \cdots, m, \tag{7}$$

where Θ is proportional to the blocker density λ_B and it is given by

$$\Theta = \frac{2}{\pi} \lambda_B V \frac{h_B - h_R}{h_T - h_R},$$

where V is the speed of the blocker, h_B, h_T and h_R are the heights of the blocker, the transmitter and the receiver, respectively. The expected blocker arrival rate is

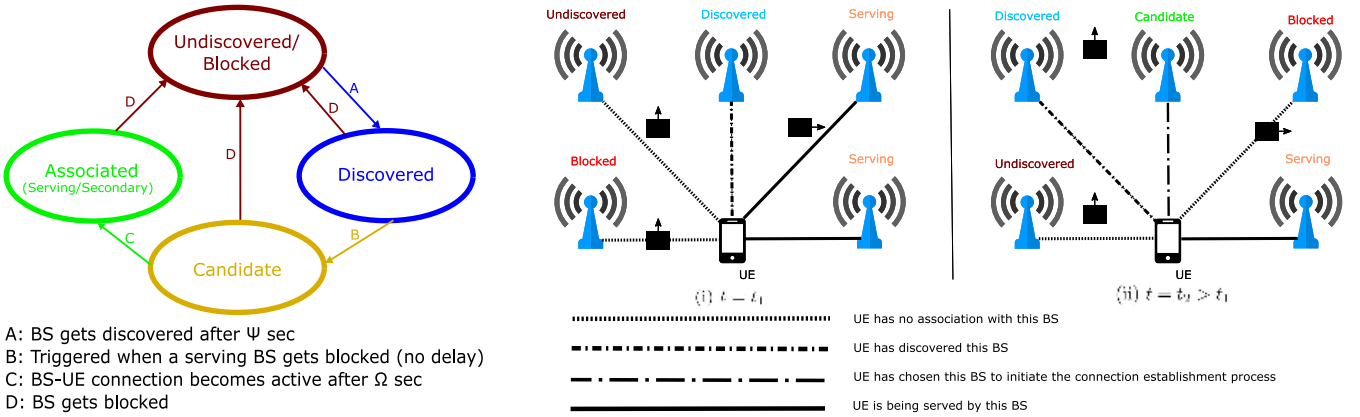
$$\mathbb{E}\left[\alpha\right] = \int_{r=0}^{R} \Theta r \frac{2r}{R^2} dr = \frac{2\Theta R}{3}. \tag{8}$$

TABLE II
SUMMARY OF NOTATIONS

Notation	Description
R	UE LOS coverage range. The UE LOS coverage range can be different at mmWave (around 100 m [32]) and THz (23 m and 56 m [42]) frequency.
$B(o,R)$	Our area of interest consisting of a disc of radius R and origin at o.
λ_T	BS density.
λ_B	Density of dynamic blockers.
α	Average arrival rate of dynamic blockers.
κ	Self-blockage angle.
p	Probability that a BS lies outside of the self-blockage zone.
$\Delta,\ \delta$	BS discover time and the corresponding rate.
$\Gamma,\ \gamma$	Blockage duration and the corresponding rate.
$\Psi,\ \psi$	Average time taken by the UE to discover a BS after BS gets blocked and the corresponding rate. Ψ includes both blockage duration and BS discovery time.
$\Omega,\ \omega$	Handover execution time and the corresponding rate.
K	Maximum degree of connectivity supported at the UE.
$P_{\mathcal{C}y}^{M\times K}$	Probability that UE is associated with $y \le K$ BSs, when M BS are in the UE coverage region and UE can achieve a maximum of K connectivity.
$P_{OS}^{M\times K}$	Out-of-service probability when M BS are in the UE coverage region and the UE can achieve a maximum of K connectivity.
$\mathcal{C}^{LOS}$	LOS Coverage considering static and self-blockage.

B. Handover Process Details

In our system, we model the handover process in conjunction with the physical blockage process in a multi-connectivity setting. The details are as follows:



(a) BS state diagram.

(b) BS-UE links status in a dual-connectivity example.

Fig. 2. (a) The BS state is determined by the BS-UE link and connection status: until a BSs-UE link becomes *unblocked*, the BS cannot be discovered by the UE. Once a BS-UE link gets unblocked, the BS remains undiscovered until the UE discovers the BS through physical layer procedures, such as cell search. Furthermore, when one of the associated BSs gets blocked, to replace it a *discovered* BS becomes a *candidate* BS for the BS-UE connection, i.e., the discovered BS is considered for the handover procedure. Furthermore, once the BS is considered for handover, the association process or equivalently the *handover execution* can take Ω time units, before the BS can be associated with the UE. (b) A dual connectivity example: A *secondary* BS in the associated list becomes the *serving* BS immediately after the link connecting the UE and the serving BS is blocked. Simultaneously, the UE initiates the association process with one of the discovered BSs.

1) Connectivity Model: In the multi-connectivity setting, we assume that the UE can be *associated* with multiple BSs at the same time. Specifically, the UE has a data and control plane connection with one *serving/master BS*, while it can maintain a data plane connection with the rest of the associated BSs, which we will call *secondary BSs*. The UE maintains a list of associated/active BSs, i.e., a list that consists of the serving/master BS and the secondary BSs. In our system, we do not distinguish among the serving/master and secondary BSs, i.e., the data plane is not considered interrupted unless and until all the serving/master and secondary BSs get blocked. The UE initiates a handover procedure to a discovered BS, once a serving/master or secondary BS gets blocked. On the initialization of the handover procedure to a discovered BS, we move the BS to the candidate set, and once the handover is completed that BS becomes an associated BS. Note that BSs in the discovered set are BSs for which the UE can obtain the Radio Resource Management (RRM) measurements [42]. If a BS is blocked, the UE will not obtain the RRM measurements for that BS. A UE is considered being out-of-service in the following scenarios: 1) out of coverage: the UE is out of coverage, i.e., no serving BSs are in the UE coverage region, 2) physically blocked: the UE is completely blocked from all of the BSs in the UE coverage region, or 3) blocked by protocol: a new BS is not added in time to the list of associated BSs before the only remaining serving BS also gets blocked, due to high BS discovery and/or handover execution times.

2) BS Status: Fig. 2(a) illustrates the transitions of the BS status. The BS status depends upon whether the link between the BS and the UE is blocked or unblocked. Until a BS-UE link becomes unblocked, the BS cannot be discovered by the UE. Even after a BS-UE link gets unblocked, it remains undiscovered until the UE discovers the BS through physical layer procedures, such as a cell search and measurement reports. Let us assume that the UE needs Δ time units to discover the unblocked BSs in its coverage region. If we also take into account that the BS-UE link remained blocked for Γ time, then the BS remains *undiscovered* by the UE for a duration of $\Psi = \Delta + \Gamma$. After the UE finds the BS through measurement reports or cell-search, the BS status changes to *discovered*. The BS remains in this status, until one of the associated BSs to the UE gets blocked. Then, one of the discovered BSs becomes a *candidate* BS for the BS-UE connection, i.e., it is considered for the handover procedure. Furthermore, once the BS is considered for handover, the association process or equivalently the handover execution can take up to Ω time units. Fig. 2(b) illustrates two snapshots of the BS-UE links status. In this example, there are five BSs in the UE coverage region, and the degree of connectivity is 2. In the first snapshot, two of the BSs are not associated with the UE: one is blocked and the other has recently been unblocked, thus it will be discovered by the UE after Δ time units. Furthermore, one BS has already been discovered by the UE and two BSs are associated with the UE. In the second snapshot (at a later time than the first snapshot), one of the associated BSs gets blocked, thus the UE begins a connection establishment process with the discovered BS, which will be completed after a handover execution duration of Ω time units.

In [14], [15], macro-diversity is considered with the assumption that a UE can switch to any of the BSs in its coverage region instantaneously, i.e., the BS discovery time, and the handover execution time are completely neglected. In [31], an alternative mmWave transport network architecture is proposed considering a similar connectivity model as discussed here. However, [31] did not incorporate the details of the handover procedure. In the next section, we will present an in-depth theoretical analysis of the handover procedure in a multi-connectivity setting that takes into account the BS discovery and handover execution times.

IV. Handover Analysis

In this section, we evaluate the impact of blockage duration Γ, BS discovery time Δ, and handover execution time Ω on the service reliability and latency of applications in the blockage

driven mmWave and THz cellular networks. Recall that, for the discovery of BSs in the UE coverage region, the BSs must be unblocked and discovered by measurement report or cell-search procedures, which takes an average of $\Psi = \Delta + \Gamma$. For tractability, we model the total BS discovery time as an exponential random variable with parameter $\psi = 1/\Psi$. We also model the handover execution time as an exponential random variable with parameter $\omega = 1/\Omega$ [43], [44]. We further assume that the blocker arrival rate at each BS-UE link follows the same exponential distribution with mean $E[\alpha]$, as calculated in (8). Using these parameters, we develop Markov chains for the 1×1 connectivity model (for ease of presentation) in Fig. 3 and for the general case $M \times K$ in Fig. 4, where M is the number of BSs in the UE coverage region and K is the degree of connectivity maintained by the UE. Note that $M \geq K$, as the multi-connectivity degree cannot exceed the number of available BSs. We denote the states of the Markov chains as $[x, y]$, where x is the number discovered BSs with an unblocked LOS link to the UE, and y is the number of associated BSs to the UE (serving and secondary).

A. 1×1 Connectivity

For the purpose of presentation ease, we will first present a continuous time Markov chain depicting the handover process in the simple 1×1 setting is shown in Fig. 3, where 1×1 implies that there is a single BS in the UE coverage region and that the UE can support at most single connectivity. In the state $[0, 0]$, the BS-UE link is blocked or this BS is currently undiscovered. In the state $[1, 0]$, the BS-UE link is discovered but the BS-UE connection has not been yet established. In the state $[1, 1]$, the BS-UE connection has been established. Thus, a UE is considered being out-of-service in the 1×1 connectivity scenario if either the BS-UE link is blocked (state $[0, 0]$) or the BS is discovered but the BS-UE connection has not been yet established (state $[1, 0]$).

Lemma 1: For the 1×1 connectivity scenario, the probability of the UE being out-of-service is

$$P_{OS}^{1\times1} = \frac{\alpha}{\alpha+\omega} + \frac{\alpha}{\alpha+\psi}\frac{\omega}{\alpha+\omega}. \tag{9}$$

Proof: To obtain the probability of the UE being out-of-service, we will solve the Markov chain in Fig. 3. Let us assume that $P_{00}^{1\times1}, P_{10}^{1\times1}$, and $P_{11}^{1\times1}$ are the steady state probabilities of states $[0, 0]$, $[1, 0]$, and $[1, 1]$, respectively, for the 1×1 connectivity scenario. Then, we have

$$\begin{aligned} \psi P_{00}^{1\times1} &= \alpha P_{10}^{1\times1} + \alpha P_{11}^{1\times1}, \\ (\alpha+\omega)\, P_{10}^{1\times1} &= \psi P_{00}^{1\times1}, \\ \alpha P_{11}^{1\times1} &= \omega P_{10}^{1\times1}. \end{aligned} \tag{10}$$

By solving the set of linear equations in (10) and $P_{00}^{1\times1} + P_{10}^{1\times1} + P_{11}^{1\times1} = 1$, we can derive the state probabilities as

$$\begin{aligned} P_{00}^{1\times1} &= \frac{\alpha}{\alpha+\psi}, \\ P_{10}^{1\times1} &= \frac{\alpha}{\alpha+\omega}\frac{\psi}{\alpha+\psi}, \\ P_{11}^{1\times1} &= \frac{\omega}{\alpha+\omega}\frac{\psi}{\alpha+\psi}. \end{aligned} \tag{11}$$

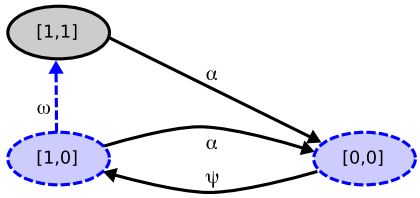


Fig. 3. Markov chain for the 1×1 connectivity model, where a single BS lies in the UE coverage region and the UE can support at maximum single connectivity. Here, in state $[0, 0]$ the BS-UE link is blocked or undiscovered, in state $[1, 0]$ the BS-UE link is unblocked, the UE has discovered the BS but the BS-UE connection has not been yet established, and in state $[1, 1]$ the BS starts serving the UE.

Using (11), the probability of the UE being out-of-service can be derived as

$$P_{OS}^{1\times1} = P_{00}^{1\times1} + P_{10}^{1\times1} = \frac{\alpha}{\alpha+\omega} + \frac{\alpha}{\alpha+\psi}\frac{\omega}{\alpha+\omega}.$$

This concludes the proof. ■

B. $M \times K$ Connectivity: Out-of-Service Probability

A Markov chain representing the generalized $M \times K$ connectivity model is presented in Fig. 4, where M BSs lie in the UE coverage region and the UE can support a maximum multi-connectivity degree of K. In the state $[x, y]$, x represents the number of unblocked BS-UE links and y represents the number of associated BSs (serving and secondary). In the state $[0, 0]$, all of the BSs are blocked or undiscovered. In the states $[1, 0], \cdots, [M, 0]$, the number of discovered BSs are $1, 2, \cdots, M$, respectively, but none of the BS-UE connections has been established yet. While being in the states $[0, 0], \cdots, [M, 0]$, the UE is considered out-of-service and experiences a data plane interruption. Furthermore, the data plane interruption continues till a transition from those states to $[1, 1], \cdots, [M, 1]$, i.e., until it regains a serving BS. Note that in all other state transitions there will be no data plane interruption, as we consider a MBB and synchronized RACH-less handover procedure during those transitions.

Proposition 1: In the Markov chain representing the generalized $M \times K$ connectivity model (see Fig. 4), the sum of all the state probabilities where $l \in \{0, 1, \cdots, M\}$ BS-UE links are unblocked is independent of K and is given by

$$P_l^{M\times K} = \sum_{i=0}^{\min(l,K)} P_{li}^{M\times K} = \binom{M}{l}\left(\frac{\psi}{\alpha+\psi}\right)^l\left(\frac{\alpha}{\alpha+\psi}\right)^{M-l}. \tag{12}$$

Proof: See Appendix A. ■

Proposition 1 implies that the number of unblocked and discovered BSs in the UE coverage region depends on the BS discovery rate and the physical blockage characteristics of the system, i.e., the blocker arrival and departure rates.

Proposition 2: If $M \leq K$ in $M \times K$ connectivity model, the probability of the UE being associated to y BSs is given by

$$\begin{aligned} P_{\mathcal{C}y}^{M\times K} &= \sum_{l=y}^{M} P_{ly}^{M\times K} = \binom{M}{y} \\ &\times \left[\frac{\alpha}{\alpha+\omega} + \frac{\alpha}{\alpha+\psi}\frac{\omega}{\alpha+\omega}\right]^M \left[\frac{\psi\omega}{\alpha(\alpha+\psi+\omega)}\right]^y \end{aligned} \tag{13}$$

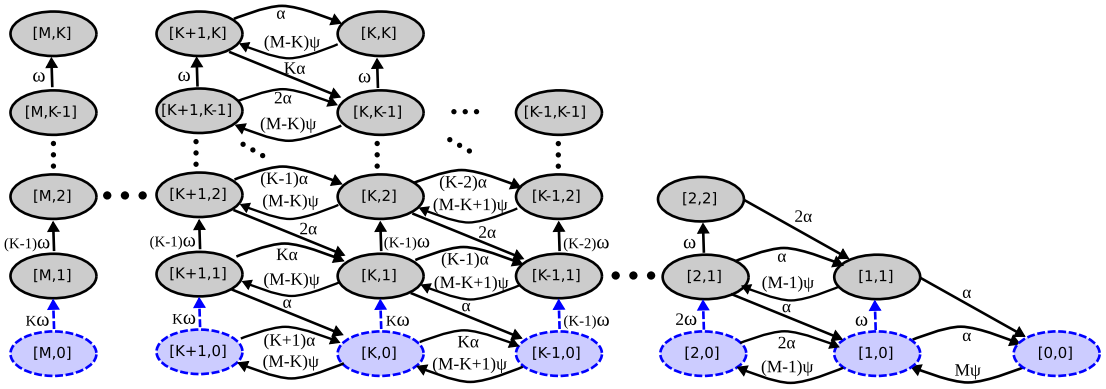


Fig. 4. Markov chain for the $M \times K$ connectivity model, where M BSs lie in the UE coverage region and the UE can support up to K connectivity. Here, in state $[0,0]$ all BS-UE links are blocked. In states $[1,0], \cdots, [M,0]$, $1, 2, \cdots, M$ BS-UE links are unblocked, respectively, but none of the BSs are in connected mode. The UE experiences data plane interruption while being in states $[0,0], \cdots, [M,0]$ and until it reaches the states $[1,1], \cdots, [M,1]$, i.e., until it stops being out-of-service. In any state $[x,y]$, x BSs are discovered with unblocked LOS links to the UE, and y BSs are associated to the UE (serving and secondary).

Proof: See Appendix B. ∎

In Corollary 1, using Proposition 2, we compute the out-of-service probability, i.e., the probability of the UE being associated with no BS, i.e., $y = 0$.

Corollary 1: For $M \leq K$, the out-of-service probability is

$$P_{OS}^{M\times K} = \left(\frac{\alpha}{\alpha+\omega} + \frac{\alpha}{\alpha+\psi}\frac{\omega}{\alpha+\omega}\right)^M. \tag{14}$$

Proof: The out-of-service probability for $M \times K, \forall M \leq K$ connectivity model can be obtained from Proposition 2 by putting $y = 0$, i.e., it can be derived by considering the scenario in which the UE is associated to none of the BSs in its coverage region. Thus, we obtain the out-of-service probability for $M \leq K$ connectivity setting as:

$$P_{OS}^{M\times K} = \left(\frac{\alpha}{\alpha+\omega} + \frac{\alpha}{\alpha+\psi}\frac{\omega}{\alpha+\omega}\right)^M, \ \forall M \leq K.$$

This concludes the proof. ∎

Note that Proposition 2 is valid only when $M \leq K$, i.e., the number of BSs in the UE coverage region is smaller than the maximum supported connectivity degree. Next, we will obtain a generalized expression for the out-of-service probability in Proposition 3.

Proposition 3: In a generalized $M \times K$ connectivity scenario, the out-of-service probability can be derived as

$$P_{OS}^{M\times K} = P_0^{M\times K} - \sum_{i=2}^{M} \min(i-1, K-1) \times P_{i0}^{M\times K} + \frac{\alpha}{\omega} P_{C1}^{M\times K}. \tag{15}$$

Proof: See Appendix C. ∎

In Corollary 2, using Proposition 3, we obtain the out-of-service probability when $K = 1$, i.e., when the UE can only achieve single-connectivity.

Corollary 2: In the $M \times 1$ connectivity setting, i.e., single connectivity, the out-of-service probability is given by

$$P_{OS}^{M\times 1} = \frac{\alpha}{\alpha+\omega} + \left(\frac{\alpha}{\alpha+\psi}\right)^M \frac{\omega}{\alpha+\omega} \tag{16}$$

Proof: In the $M \times 1$ connectivity setting, from Proposition 3, we have

$$\begin{aligned} P_{OS}^{M\times 1} &= P_0^{M\times 1} + \frac{\alpha}{\omega} P_{C1}^{M\times 1} \\ &= P_0^{M\times 1} + \frac{\alpha}{\omega}\left(1 - P_{OS}^{M\times 1}\right) \\ &= \frac{\alpha}{\alpha+\omega} + \frac{\omega}{\omega+\alpha} P_0^{M\times 1}. \end{aligned} \tag{17}$$

Furthermore, using Proposition 1 and (17), we obtain the out-of service probability in the $M \times 1$ connectivity setting as

$$P_{OS}^{M\times 1} = \frac{\alpha}{\alpha+\omega} + \left(\frac{\alpha}{\alpha+\psi}\right)^M \frac{\omega}{\alpha+\omega}$$

This concludes the proof. ∎

Note that in order to compute the second term in Proposition 1, we need the individual state probabilities of the generalized $M \times K$ Markov chain. Thus, we will instead use Proposition 1 to obtain upper and lower bounds on the out-of-service probability for the generalized $M \times K$ scenario.

Corollary 3: The upper and lower bounds on the out-of-service probability for $M \times K$ are given by

$$\begin{aligned} P_{OS,UB}^{M\times K} &\leq P_0^{M\times K} + \frac{1}{2} P_{10}^{K\times K} + \frac{\alpha}{2\omega} P_{C1}^{K\times K} \\ &= \left[\frac{\alpha}{\alpha+\psi}\right]^M + \frac{K}{2}\left[\frac{\alpha}{\alpha+\psi}\right]^K \left[\frac{\psi}{\alpha+\omega}\right] \\ &\quad + \frac{K\alpha}{2\omega}\left[\frac{\alpha}{\alpha+\omega} + \frac{\alpha}{\alpha+\psi}\frac{\omega}{\alpha+\omega}\right]^{K-1} \\ &\quad \times \left[\frac{\psi\omega}{(\alpha+\psi)(\alpha+\omega)}\right] \end{aligned} \tag{18}$$

$$\begin{aligned} P_{OS,LB}^{M\times K} &\geq P_0^{M\times K} + \frac{K-1}{K} P_{10}^{M\times M} + \frac{\alpha}{K\omega} P_{C1}^{M\times M} \\ &= \left[\frac{\alpha}{\alpha+\psi}\right]^M + M\frac{K-1}{K}\left[\frac{\alpha}{\alpha+\psi}\right]^M \left[\frac{\psi}{\alpha+\omega}\right] \\ &\quad + \frac{M\alpha}{K\omega}\left[\frac{\alpha}{\alpha+\omega} + \frac{\alpha}{\alpha+\psi}\frac{\omega}{\alpha+\omega}\right]^{M-1} \\ &\quad \times \left[\frac{\psi\omega}{(\alpha+\psi)(\alpha+\omega)}\right] \end{aligned} \tag{19}$$

Proof: See Appendix D. ∎

Using Corollaries 1-3, we derive the out-of-service probability given m BSs in the UE coverage region in (20), as shown at the bottom of the page. Finally, using (5), (6), and (20), we derive the out-of-service probability given LOS coverage in Theorem 1.

Theorem 1: We obtained the exact expression for out-of-service probability given LOS coverage for single connectivity, and lower and upper bounds on the out-of-service probability given LOS coverage for connectivity $K \geq 2$. The out-of-service probability given LOS coverage is given in (21)*, as shown at the bottom of the page.*

Proof: See Appendix E. ∎

Corollary 4 is obtained using Theorem 1, for the case when the handover execution rate is significantly higher than the blockage arrival and the BS discovery rates. In such scenario, the out-of-service probability becomes independent of the degree of multi-connectivity. Thus, in cellular transport network architectures, if the handover latency can be reduced significantly, then a high degree of connectivity will be not be required to meet the QoS requirements.

Note that the equation in Corollary 4 is the same as the result for the blockage probability given LOS coverage assuming instantaneous handover among BSs in [14], [15], which shows that if the handover execution and BS discovery times are close to zero, we obtain the same result.

Corollary 4: If $\omega \gg \alpha$ and $\omega \gg \psi$, the out-of-service probability can be approximated by

$$P\left(OS|C^{LOS}\right) = \frac{e^{-qp\lambda_T\pi R^2} - e^{-p\lambda_T\pi R^2}}{1 - e^{-p\lambda_T\pi R^2}}. \tag{22}$$

Proof: See Appendix F. ∎

In Corollary 5, using Corollary 4, we try to answer what BS discovery time will be sufficient to achieve a certain out-of-service probability, $P\left(OS|C^{\text{LOS}}\right)$, in the blockage driven THz and mmWave cellular networks. Thus, Corollary 5 can be used to study the trade-off between maximum allowed BS discovery time and minimum BS density required to achieve the desired reliability.

Corollary 5: If $\omega \gg \alpha$ and $\omega \gg \psi$, then the maximum allowed BS discovery time to achieve a certain out-of-service probability, $P\left(OS|C^{LOS}\right)$, is given by

$$\Delta = -\frac{(1+\Upsilon)}{\Upsilon\alpha} - \Gamma,$$

where ,

$$\Upsilon = \frac{ln\left[P\left(OS|C^{LOS}\right) + e^{-p\lambda_T\pi R^2}\left(1 - P\left(OS|C^{LOS}\right)\right)\right]}{p\lambda_T\pi R^2}. \tag{23}$$

Proof: See Appendix G. ∎

$$\begin{aligned}
P\left(OS|m\right) &= \frac{\alpha}{\alpha+\omega} + \left(\frac{\alpha}{\alpha+\psi}\right)^M \frac{\omega}{\alpha+\omega}; \text{if } K=1, \forall m \in \{1,2,\cdots,M\},\\
&= \left[\frac{\alpha}{\alpha+\omega} + \frac{\alpha}{\alpha+\psi}\frac{\omega}{\alpha+\omega}\right]^m; \text{if } K \geq 2, \forall m \in \{1,2,\cdots,K\},\\
&\leq \left[\frac{\alpha}{\alpha+\psi}\right]^m + \frac{K}{2}\left[\frac{\alpha}{\alpha+\psi}\right]^K\left[\frac{\psi}{\alpha+\omega}\right] + \frac{K}{2}\left[\frac{\alpha}{\alpha+\omega} + \frac{\alpha}{\alpha+\psi}\frac{\omega}{\alpha+\omega}\right]^{K-1}\left[\frac{\psi}{\alpha+\psi}\frac{\alpha}{\alpha+\omega}\right];\\
&\quad \text{if } K \geq 2, \forall m \in \{K+1, K+2, \cdots\},\\
&\geq \left[\frac{\alpha}{\alpha+\psi}\right]^m + m\frac{K-1}{K}\left[\frac{\alpha}{\alpha+\psi}\right]^m\left[\frac{\psi}{\alpha+\omega}\right] + \frac{m}{K}\left[\frac{\alpha}{\alpha+\omega} + \frac{\alpha}{\alpha+\psi}\frac{\omega}{\alpha+\omega}\right]^{m-1}\left[\frac{\psi}{\alpha+\psi}\frac{\alpha}{\alpha+\omega}\right];\\
&\quad \text{if } K \geq 2, \forall m \in \{K+1, K+2, \cdots\}.
\end{aligned} \tag{20}$$

$$\begin{aligned}
P\left(OS|C^{\text{LOS}}\right) &= \frac{1 - \frac{\omega}{\alpha+\omega}\left(1 - e^{-qp\lambda_T\pi R^2}\right) - e^{-p\lambda_T\pi R^2}}{1 - e^{-p\lambda_T\pi R^2}}; \text{if } K=1,\\
&\leq \frac{e^{-p\lambda_T\pi R^2}}{1 - e^{-p\lambda_T\pi R^2}}\left[\sum_{m=1}^{K}\frac{\left[cp\lambda_T\pi R^2\right]^m}{m!} + \sum_{m=K+1}^{\infty}\left(\frac{\left[\tilde{q}p\lambda_T\pi R^2\right]^m}{m!} + \chi\frac{\left[p\lambda_T\pi R^2\right]^m}{m!}\right)\right]; \text{if } K \geq 2,\\
&\geq \frac{e^{-p\lambda_T\pi R^2}}{1 - e^{-p\lambda_T\pi R^2}}\left[\sum_{m=1}^{K}\frac{\left[cp\lambda_T\pi R^2\right]^m}{m!} + \sum_{m=K+1}^{\infty}\left[1 + \frac{m\psi(K-1)}{K(\alpha+\omega)}\right]\frac{\left[\tilde{q}p\lambda_T\pi R^2\right]^m}{m!} + \frac{\alpha\zeta}{K\omega}\right.\\
&\quad \left.\sum_{m=K+1}^{\infty}\frac{\left[cp\lambda_T\pi R^2\right]^m}{(m-1)!}\right]; \quad \text{if } K \geq 2,\\
\text{where, } c &= P_{OS}^{1\times 1},\ q = \frac{\psi}{\alpha+\psi},\ \tilde{q} = (1-q), \chi = \frac{K}{2}\left[\frac{\alpha}{\alpha+\psi}\right]^K\left[\frac{\psi}{\alpha+\omega}\right]\\
&\quad + \frac{K}{2}\left[\frac{\alpha}{\alpha+\omega} + \frac{\alpha}{\alpha+\psi}\frac{\omega}{\alpha+\omega}\right]^{K-1}\left[\frac{\psi}{\alpha+\psi}\frac{\alpha}{\alpha+\omega}\right],\\
\text{and } \zeta &= \frac{\psi\omega}{\alpha(\alpha+\psi+\omega)}.
\end{aligned} \tag{21}$$

Corollary 6 is obtained using Theorem 1, when the UE is not limited by the degree of connectivity, i.e., the UE has sufficient RF chains and antenna elements available to exploit BS available in its overage region. In such a scenario, the out-of-service probability is limited by how efficiently UE can exploit the discovered BSs in its coverage region.

Corollary 6: When $K \to \infty$, i.e., the UE can connect to all of the BSs in its coverage region if they are not blocked, then the out-of-service probability given LOS coverage is given by

$$\lim_{K\to\infty} P\left(OS|C^{LOS}\right) = \frac{e^{-\tilde{c}p\lambda_T\pi R^2} - e^{-p\lambda_T\pi R^2}}{1-e^{-p\lambda_T\pi R^2}} \quad \text{where, } \tilde{c} = 1 - P_{OS}^{1\times 1}. \tag{24}$$

Proof: See Appendix H. ∎

We can use Corollary 6 to examine whether the reliability requirement of different applications and services can be achieved for a choice of handover execution and BS discovery times given a sufficiently high degree of connectivity.

C. $M \times K$ Connectivity: RLF Probability

An RLF is detected by the UE when it is out-of-service for a determined time duration. In particular, RLF is declared upon the expiration of the T310 timer [45]. In the Markov chain presented in Fig. 4, RLF will be declared if there is a transition from $[i,1]$, to $[i-1,0],\ \forall i \in \{1,\cdots,M\}$ and the time to transition back to $[j,1],\ j \geq 1$ exceeds $\mathcal{T}$ time units. If there is a transition back to $[j,1],\ j \geq 1$, within $\mathcal{T}$ time units, the T310 timer stops counting down and an RLF is avoided. On the other hand, if there is a transition from $[i,0]$ to $[j,0],\ j \neq i$ within $\mathcal{T}$, the T310 timer will not reset and continue to count down. However, for analytical tractability of the RLF probability, we assume, that once the T310 timer starts counting down when entering $[i,0]$, the first state to transition will be $[i,1], \forall i \in \{1,\cdots,M\}$ i.e., we do not take into account transitions from $[i,0]$ to $[j,0],\ j \neq i$ within $\mathcal{T}$. Note that since we exclude transitions to neighboring states $[j,0],\ j \neq i$ within $\mathcal{T}$, we decrease the time spent in the out-of-service states, which results in a lower bound on the RLF probability. Indeed, we argue that for the range of parameters we are interested in, the transition rate from [i,0] to [i,1] (around 50 - 100) is significantly larger than the transition rates to [i-1,0] (around 0.05 - 0.5) and [i+1,0] (around 1- 2), thus we obtain a tight lower bound on the RLF probability. This lower bound is given by

$$\begin{aligned} P_{RLF}^{M\times K} &\geq P_{00}^{M\times K} P\left(\exp(M\psi) > \mathcal{T}\right) \\ &\quad + \sum_{m=1}^{M} P_{m0}^{M\times K} P\left(\exp(\eta_m\omega) > \mathcal{T}\right) \\ &\geq P_{0S}^{M\times K} P\left(\exp(K\omega) > \mathcal{T}\right) \\ &\quad + P_{00}^{M\times K}\left(P\left(\exp(M\psi) > \mathcal{T}\right)\right. \\ &\quad \left. - P\left(\exp(K\omega) > \mathcal{T}\right)\right) \\ &= P_{0S}^{M\times K} e^{-K\omega\mathcal{T}} + P_{00}^{M\times K}\left(e^{-M\psi\mathcal{T}} - e^{-K\omega\mathcal{T}}\right) \\ &\quad \text{where, } \eta_m = \min(m,K). \end{aligned} \tag{25}$$

Theorem 2: A lower bound on the RLF probability given LOS coverage is derived as

$$\begin{aligned} &P\left(RLF|C^{LOS}\right) \\ &\geq P\left(OS|C^{LOS}\right) e^{-K\omega\mathcal{T}} + \frac{e^{-p\lambda_T\pi R^2}}{1-e^{-p\lambda_T\pi R^2}} \\ &\quad \times \left[e^{\xi\tilde{q}p\lambda_T\pi R^2} - e^{-K\omega T}\left(e^{\tilde{q}p\lambda_T\pi R^2} - 1\right) - 1\right], \end{aligned}$$

where , $\xi = e^{-\psi\mathcal{T}}$ *and* $\tilde{q} = 1 - q = \dfrac{\alpha}{\alpha+\psi}$. (26)

Proof: See Appendix I. ∎

Corollary 7 is obtained using Theorem 2 when the handover execution rate is significantly higher than the blockage arrival and the BS discovery rates.

Corollary 7: If $\omega \gg \alpha$ and $\omega \gg \psi$, then the RLF probability can be reduced to

$$P\left(RLF|C^{LOS}\right) \geq \frac{e^{-(1-\xi\tilde{q})p\lambda_T\pi R^2} - e^{-p\lambda_T\pi R^2}}{1-e^{-p\lambda_T\pi R^2}}. \tag{27}$$

Proof: See Appendix J. ∎

We further develop Corollary 8 using Corollary 7 to present the trade-off between the maximum allowed BS discovery time and the minimum required BS density for meeting a given RLF probability, $P\left(RLF|C^{\text{LOS}}\right)$.

Corollary 8: If $\omega \gg \alpha$ and $\omega \gg \psi$, then the maximum allowed BS discovery time to achieve a certain RLF probability, $P\left(RLF|C^{LOS}\right)$, is given by

$$\Delta = \frac{\mathcal{T}}{W_n\left(\frac{\alpha\mathcal{T}e^{\alpha\mathcal{T}}}{\eta+1}\right) - \alpha\mathcal{T}} - \Gamma,$$

where,

$$\eta = \frac{ln\left[P\left(RLF|C^{LOS}\right) + e^{-p\lambda_T\pi R^2}\left(1 - P\left(RLF|C^{LOS}\right)\right)\right]}{p\lambda_T\pi R^2}, \tag{28}$$

and $W_k(z)$ is the analytic computation of the product log function.

Proof: See Appendix K. ∎

Corollary 9 is obtained using Theorem 1 and Theorem 2 when the handover execution rate is significantly higher than the blockage arrival and the BS discovery rates.

Corollary 9: If $\omega \gg \alpha$ and $\omega \gg \psi$, and $\delta \gg \gamma$, then the ratio of the RLF probability to the out-of-service probability has the following lower bound

$$\frac{P\left(RLF|C^{LOS}\right)}{P\left(OS|C^{LOS}\right)} \geq \frac{e^{\xi\tilde{q}p\lambda_T\pi R^2} - 1}{e^{\tilde{q}p\lambda_T\pi R^2} - 1} \approx \left(e^{-\tilde{q}p\lambda_T\pi R^2}\right)^{1-\xi} \tag{29}$$

Proof: See Appendix L. ∎

From Corollary 7 and Corollary 9, we can observe that most of the out-of-service instances will result in RLF failure if $\mathcal{T} \to 0$. On the other hand, note that $\mathcal{T}$ can be as large as 2 seconds [42]. Thus, if $\mathcal{T}$ is significantly large, then $\xi \to 0$ and the RLF failure rate will in turn reduce to $e^{-\tilde{q}p\lambda_T\pi R^2}$. However, if $\mathcal{T} \to \infty$, the UE will have to suffer a very large out-of-service duration for an RLF to be declared. Moreover, for a small mean blockage duration, or equivalently when $\psi \to \infty$, then $\xi \to 0$, which will greatly reduce the RLF probability.

D. $M \times K$ Connectivity: Expected Out-of-Service Duration

In the Markov chain of Fig. 4, the data plane connection is interrupted when there is a transition from any connected state to an out-of-service state, i.e., a state $[k,0], \forall k \in \{0,1,\cdots,M\}$. Furthermore, data communication remains suspended until there is a transition from the out-of-service states to the connected states, i.e., it enters $[k,1], \forall k \in \{1,\cdots,M\}$. Note from Fig. 4 that there can be direct transitions from $[k,0]$ only to $[k,1]$, $[k+1,0]$, or $[k-1,0]$. Thus, the average out-of-service duration $\mathbf{t}^{\mathbf{M\times K}}([k,0])$, for state $[k,0]$ can be derived as the weighted sum of

1) time to transition from $[k,0]$ to $[k,1]$,
2) time to transition from $[k,0]$ to $[k+1,0]$ and thereafter to a connected state, and
3) time to transition from $[k,0]$ to $[k-1,0]$ and thereafter to a connected state.

In a continuous Markov chain with exponential transition rates, the expected time to transition from one state to another is an exponential random variable with rate equal to the sum of the outgoing rates [46]. Thus, for the Markov chain shown in Fig. 4, the expected duration to leave the state $[m,k]$, $m \in \{0,1,\ldots,M\},\ k \in \{0,1,\ldots,K\}$ is:

$$\mathbf{t}^{\mathbf{M\times K}}_{\text{leave}}([m,k]) = \frac{1}{(\eta_m - k)\omega + (M-m)\psi + m\alpha}, \quad \text{where } \eta_m = \min(K,m) \tag{30}$$

Thus, we derive the out-of-service duration given M BSs in the UE coverage region outside of the self-blockage zone by solving the following system of $M+1$ linear equations where we derive the duration it takes to regain service starting from state $[k,0]$, $\mathbf{t}^{\mathbf{M\times K}}([k,0])$, as:

$$\begin{aligned}
\mathbf{t}^{\mathbf{M\times K}}([m,0]) &= \mathbf{t}^{\mathbf{M\times K}}_{\text{leave}}([m,0]) \\
&\quad \times \frac{\eta_m\omega}{\eta_m\omega + (M-m)\psi + m\alpha} \\
&\quad + \left[\mathbf{t}^{\mathbf{M\times K}}_{\text{leave}}([m,0]) + \mathbf{t}^{\mathbf{M\times K}}([m-1,0])\right] \\
&\quad \times \frac{m\alpha}{\eta_m\omega + (M-m)\psi + m\alpha} \\
&\quad + \left[\mathbf{t}^{\mathbf{M\times K}}_{\text{leave}}([m,0]) + \mathbf{t}^{\mathbf{M\times K}}([m+1,0])\right] \\
&\quad \times \frac{(M-m)\psi}{\eta_m\omega + (M-m)\psi + m\alpha} \\
&= \mathbf{t}^{\mathbf{M\times K}}_{\text{leave}}([m,0]) + \mathbf{t}^{\mathbf{M\times K}}([m-1,0]) \\
&\quad \times \frac{k\alpha}{\eta_m\omega + (M-m)\psi + m\alpha} + \mathbf{t}^{\mathbf{M\times K}} \\
&\quad \times ([m+1,0]) \times \frac{(M-m)\psi}{\eta_m\omega + (M-m)\psi + m\alpha}, \\
&\quad \forall m \in \{0,\cdots,M\}, \\
&\quad \text{where } \mathbf{t}^{\mathbf{M\times K}}([M+1,0]) = 0, \\
&\quad \mathbf{t}^{\mathbf{M\times K}}([-1,0]) = 0.
\end{aligned} \tag{31}$$

We can obtain $\mathbf{t}^{\mathbf{M\times K}}([k,0]), \forall m \in \{0,1,\cdots,M\}$ by solving (31). Furthermore, the out-of-service duration $T_{OS}^{M\times K}$ when the UE is in out-of-service can be derived as (see Appendix M for details):

$$T_{OS}^{M\times K} = \frac{\sum_{m=0}^{M-1} \mathbf{t}^{\mathbf{M\times K}}_{\text{leave}}([m+1,1])\, \mathbf{t}^{\mathbf{M\times K}}([m,0])\, P_{m+1,1}^{M\times K}}{\sum_{j=1}^{M} \mathbf{t}^{\mathbf{M\times K}}_{\text{leave}}([j,1])\, P_{j1}^{M\times K}} \tag{32}$$

In order to compute the out-of-service duration from (32), we need to know each individual state probability $P_{m+1,1}^{M\times K}$, $\forall m \in \{0,\cdots,M-1\}$. This can be computed, but it is more convenient to have an expression that is simple to compute, thus, we will instead obtain a lower bound on the out-of-service duration by using Little's Law [47].

Lemma 2: In the $M \times K$ connectivity setting, the expected out-of-service duration obtained using Little's Law is given by

$$T_{OS}^{M\times K} = \frac{P_{OS}^{M\times K}}{\alpha P_{C1}^{M\times K}}. \tag{33}$$

Proof: See Appendix N. ∎

Theorem 3: The expected out-of-service duration given LOS coverage is given in (34)*, as shown at the bottom of the page.*

Proof: See Appendix O. ∎

To study the trade-off between the maximum allowed BS discovery time and minimum required BS density for meeting the out-of-service duration for different applications and services, we develop Corollary 10 using Theorem 3. We assume in Corollary 10 that the UE has a sufficiently high number of antenna elements available.

Corollary 10: When $K \to \infty$, i.e., the UE can connect to all of the BSs in its coverage region if they are not blocked, then the maximum allowed BS discovery time to achieve a certain out-of-service duration is given by

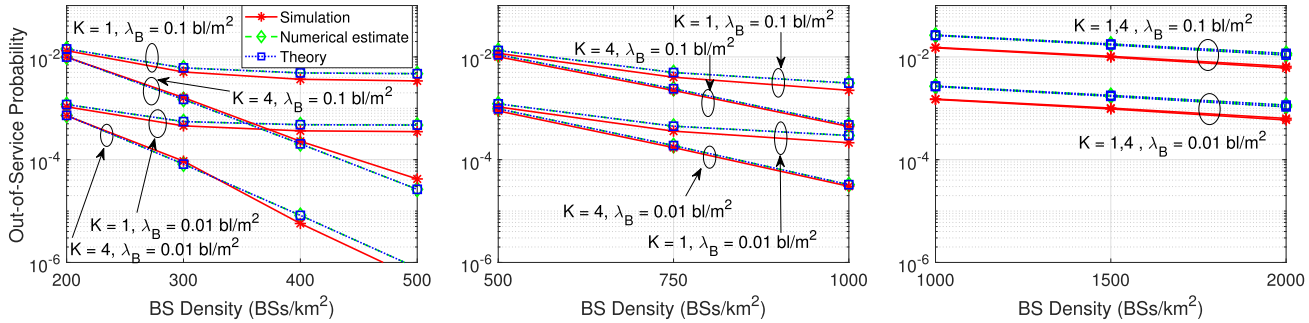
$$\Delta = \frac{\nu\omega - 1}{\alpha + \omega} - \Gamma, \quad \textit{where, } \nu = \frac{\mathbb{E}\left[T\left(OS|C^{LOS}\right)\right]\left(1 - e^{-p\lambda_T\pi R^2}\right)}{Ei\left[p\lambda_T\pi R^2\right] e^{-p\lambda_T\pi R^2}},$$
$$\textit{and } Ei\left[p\lambda_T\pi R^2\right] = \sum_{m=1}^{\infty} \frac{[p\lambda_T\pi R^2]^m}{mm!}. \tag{35}$$

Proof: See Appendix P. ∎

V. Numerical Results

For the MATLAB simulation, we considered a square of size $2R \times 2R$ with blockers located uniformly in this area, where R is the radius of disc $B(o,R)$ that perfectly fits in the considered square area. This disc represents the UE

$$\mathbb{E}\left[T\left(OS|C^{\text{LOS}}\right)\right] = \frac{\frac{1}{\omega}\left[1 - e^{-p\lambda_T\pi R^2}\right] + \frac{1}{\alpha}\left[e^{-(1-\tilde{q})p\lambda_T\pi R^2} - e^{-p\lambda_T\pi R^2}\right]}{1 - e^{-p\lambda_T\pi R^2}};\ \text{if } K = 1, \approx \frac{e^{-p\lambda_T\pi R^2}}{1 - e^{-p\lambda_T\pi R^2}} \times \left[\frac{\alpha + \psi + \omega}{\psi\omega}\sum_{m=1}^{K}\frac{(p\lambda_T\pi R^2)^m}{mm!} + \frac{1}{K\omega}\sum_{m=K+1}^{\infty}\left(1 + \frac{\tilde{q}^m}{\chi}\right)\frac{(p\lambda_T\pi R^2)^m}{m!}\right];\ \text{if } K \geq 2. \tag{34}$$



(a) mmWave: $R = 100$ m. (b) THz: $R = 56$ m. (c) THz: $R = 23$ m.

Fig. 5. Out-of-service probability in different coverage ranges (R) for mmWave and THz cellular networks: a comparison between the simulation, the numerical solution and the theoretical lower bound for different BS and blocker density values (λ_T, λ_B), different degrees of connectivity (K), for fixed BS discovery time, Δ, of 20 ms, and handover execution time, Ω, of 10 ms. Note that, theoretical bounds, numerical estimate, and simulation results agree with each other for various parameter settings.

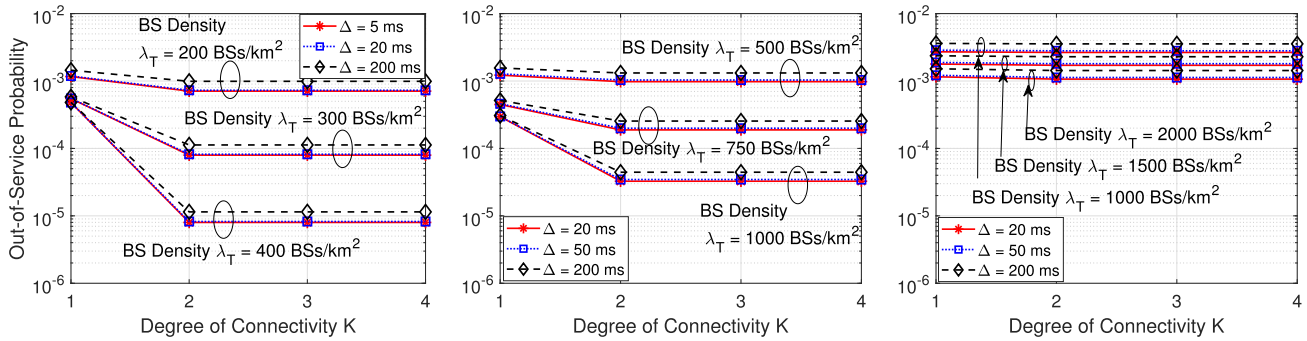
coverage region in the mmWave or THz cellular network. Note that the coverage region can be significantly smaller in THz networks as compared to legacy microwave and mmWave cellular networks. In mmWave cellular networks we assume a UE coverage range of 100 m [31], whereas in the THz cellular network we assume a UE coverage range of 23 m and 56 m [41], due to the significantly higher path loss at THz frequencies. Furthermore, as a larger number of antenna elements at the UE and BS are required for efficient beamforming to overcome the path loss in THz systems, the BS discovery time in THz systems can be significantly higher as compared to mmWave systems [20], [21]. While the BS discovery time can be reduced using digital beamforming, it requires a large number of independent RF chains making it impractical for THz systems [48]. In our simulation, we consider the BS discovery times of 5 ms and 20 ms [32] in the mmWave cellular networks, and we assume that at least an order of magnitude increase in BS discovery time will be required in THz systems. Thus, we consider the cell search time for THz networks as 20 ms, 50 ms, and 200 ms. Note that THz networks are still in an early study phase and many factors of the system design are yet to be determined. Therefore, our selection of BS discovery time in THz networks is considered just for illustration. However, *as THz systems evolve, one can use our theoretical framework to obtain system performance metrics by updating the parameter selection.*

Note that a significantly higher BS density is required in THz cellular networks due to the lower coverage range as compared to mmWave cellular networks. Thus, we consider a BS density from 500 BSs/km^2 to 2000 BSs/km^2 for the THz cellular network, whereas for the mmWave network we consider a BS density from 200 BSs/km^2 to 500 BSs/km^2. We use two dynamic blocker density values, 0.01 bl/m^2 and 0.1 bl/m^2 [15]. For blocker movement in the considered square region, we use the random waypoint model [49], [50], where blockers choose a direction randomly, and move in that direction for a time duration chosen uniformly over $(0, 60]$ seconds. We performed 10,000 runs where each run consisted of the equivalent of 4 hours of blocker mobility. To maintain a fixed density of blockers in the square region, we consider that once a blocker reaches the edge of the square, it gets reflected. The average blockage duration is 0.5 seconds [51]. Furthermore, we limit the highest achievable connectivity degree to 4, i.e., the UE can connect up to 4 BSs simultaneously if there are at least 4 BSs in the UE coverage region. Note that if there are less than 4 BSs in the UE coverage region, the UE can connect only up to the number of the BSs in the UE coverage region (0, 1, 2 or 3).

TABLE III
SIMULATION PARAMETERS

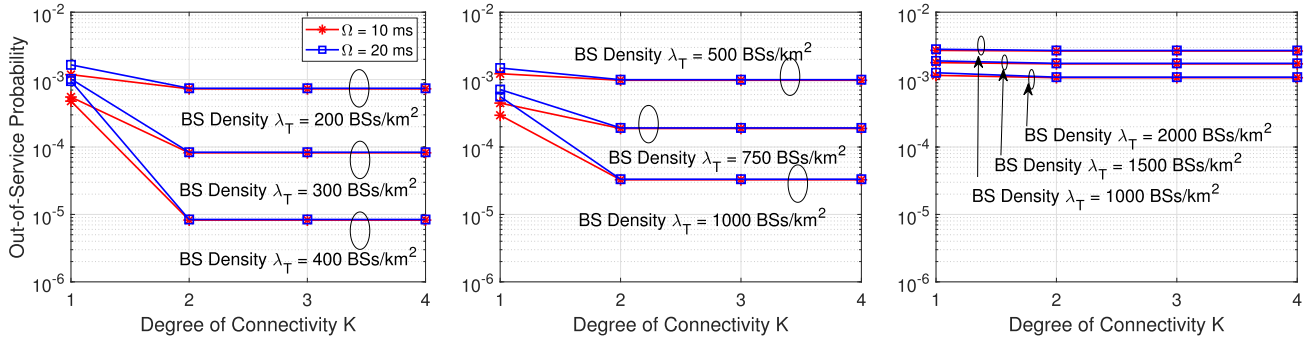
Parameters	**Values**	
LOS Radius, R	mmWave	THz
	100 m	{23, 56} m
Velocity of Dynamic Blockers, V	1 m/s	
Height of Dynamic Blockers, h_B	1.8 m	
Height of UE, h_R	1.4 m	
Height of gNB-DU, h_T	5 m	
Expected blockage duration, Γ	500 ms	
Expected BS discovery time, Δ	mmWave	THz
	{5, 20} ms	{20, 50, 200} ms
Expected handover execution time, Ω	{10, 20, 50} ms	
T310 timer, $\mathcal{T}$	50 ms	
Dynamic blocker density, λ_B	{0.01, 0.1} bl/m^2	
Base station density, λ_T [BSs/km^2]	mmWave	THz
	{200,...,500}	{500,...,2000}

In the multi-connectivity scenario, we assumed that switching to a secondary BS occurs immediately after the serving BS gets blocked. However, adding a new BS to the list of serving and secondary BSs can take up to the sum of the BS discovery and handover execution time. Thus, the UE will be out-of-service in following scenarios: 1) UE is out of coverage, i.e., there are no serving BSs in the UE coverage region, 2) the UE is completely blocked from all of the BSs in its coverage region, and 3) the only remaining serving BS gets blocked, and an unblocked BS is not added promptly enough due to BS discovery and handover execution times to prevent a period of blockage. We will refer to the second scenario as *physically blocked* and to the third scenario as *blocked by protocol*. The rest of the simulation parameters are presented in Table III.



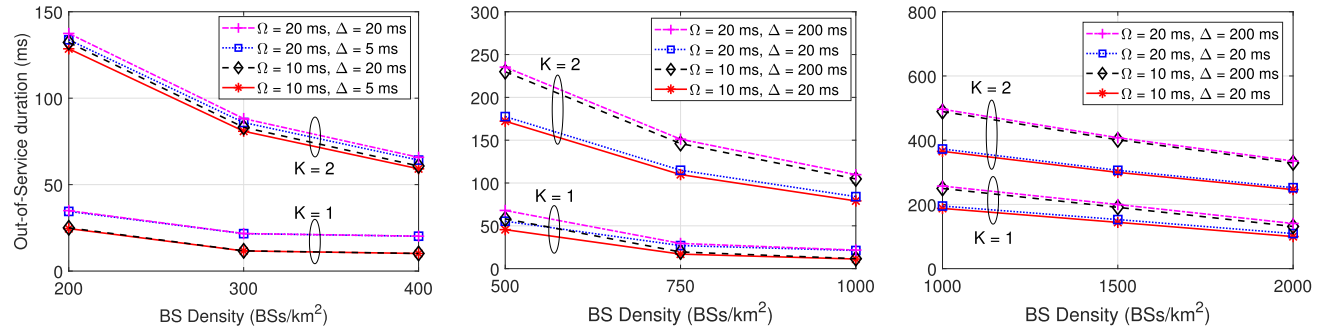
(a) mmWave: $R = 100$ m. (b) THz: $R = 56$ m. (c) THZ: $R = 23$ m.

Fig. 6. Out-of-service probability in mmWave and THz cellular networks with different degrees of connectivity K, UE coverage range R, BS density values λ_T, and BS discovery times Δ for a fixed value of handover execution time, Ω, of 10 ms, and blocker density, λ_B, of 0.01 bl/m^2. In the THz networks, the BS discovery time can be significantly higher than mmWave cellular networks due to longer time required searching for suitable beams.



(a) mmWave: $R = 100$ m. (b) THz: $R = 56$ m. (c) THz: $R = 23$ m.

Fig. 7. Out-of-service probability with different degrees of connectivity K, BS density values λ_T, and handover execution times Ω for a fixed value of BS discovery time, Δ, of 20 ms and blocker density, λ_B, of 0.01 bl/m^2.



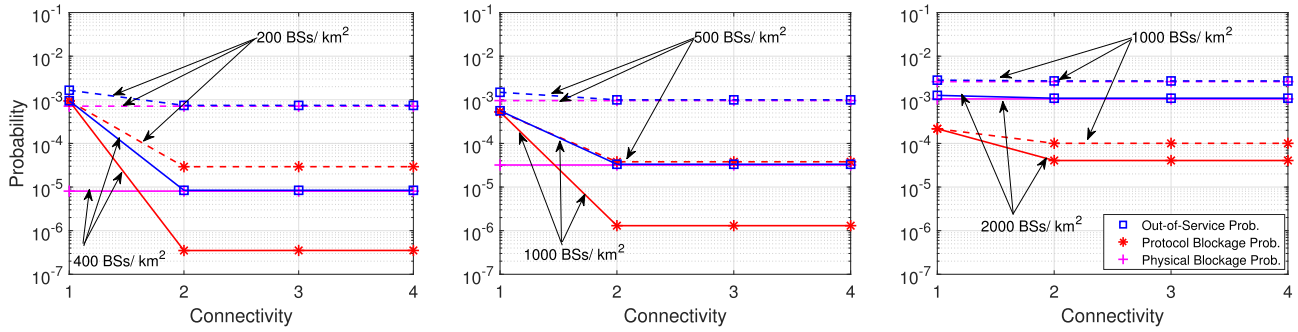
(a) mmWave: $R = 100$ m. (b) THz: $R = 56$ m. (c) THz: $R = 23$ m.

Fig. 8. Out-of-service duration: an illustration of the impact of the multi-connectivity degree K, BS density λ_T, BS discovery time Δ, and handover execution time Ω, on the expected out-of-service duration of mmWave and THz systems.

A. Out-of-Service Probability

Fig. 5, Fig. 6, and Fig. 7 plot the out-of-service probability for different values of a) coverage range in mmWave and THz cellular networks b) BS density, c) blocker density, d) degree of connectivity, and e) BS discovery and handover execution times. Fig. 5 demonstrates that our simulation and analytical model produce similar results. Fig. 6 and Fig. 7 give useful insights into the trade-off of different parameters to meet the reliability requirement of URLLC applications.

Fig. 5 illustrates a comparison of the out-of-service probability results generated by our simulation, a numerical computation of the analytical solution and the theoretical lower bounds. From Fig. 5, we observe that a lower out-of-service probability can be achieved in a mmWave network as compared to THz networks, due to its larger coverage range. Furthermore, we can observe that in the THz system if the coverage range is small, for example 23 m, even increasing the connectivity degree does not help in reducing the out-of-service probability. This happens because with low coverage range and BS density, the number of BSs that the UE can connect to is significantly reduced. Thus, to achieve a low out-of-service probability in the THz network, we may need a significantly higher BS density as compared to the mmWave network. Finally, as expected, we observe that the out-of-service probability reduces with smaller blocker density.



(a) mmWave: $R = 100$ m. (b) THz: $R = 56$ m. (c) THz: $R = 23$ m.

Fig. 9. Physical blockage probability and blockage-by-protocol probability for different multi-connectivity degrees and BS density values with $\Omega = 20$ ms and $\Delta = 20$ ms.

Note from Fig. 5 that theoretical bounds, numerical estimates, and simulation results agree with each other for various parameter settings considered for mmWave and THz cellular networks. Thus to improve readability, for the rest of the paper we will use only the easily computable theoretical bounds. Note that the small differences between the simulation and the theoretical results can be explained by considering the fact that there are two main differences between the two approaches: In our analysis, we assume the blocker arrival rate α_i of a single UE-BS link can be approximated by an exponential distribution with mean equal to $E[\alpha]$, as calculated in (8), whereas in our simulation, the blocker arrival rate (calculated using (7)) is a function of the BS-UE distance. Furthermore, in our analytical model, we assume the BS discovery and handover execution times are exponential random variables. In our simulation, the BS discovery and handover execution times are constant values, equal to the mean of the respective exponential random variables.

From Fig. 6, we observe that a lower BS discovery time helps in achieving a modestly lower out-of service probability in both mmWave and THz networks irrespective of the coverage range. Furthermore, we observe from Fig. 6(a) and Fig. 6(b) that the out-of-service probability drops sharply when connectivity increases from 1 to 2, especially for higher BS density values, for both mmWave and THz networks provided that a suitable coverage range can be achieved. However, any further increase in connectivity leads to only a marginal reduction in the out-of-service probability. Note from Fig. 6(c) and Fig. 7(c) that even increasing the UE connectivity degree beyond one does not help, due to the low number of BSs available to a typical UE for this range and BS densities. Furthermore, we can infer from Fig. 7 that in a single connectivity scenario reducing the handover execution time is much more important to obtain a lower out-of-service probability than in multi-connectivity scenarios.

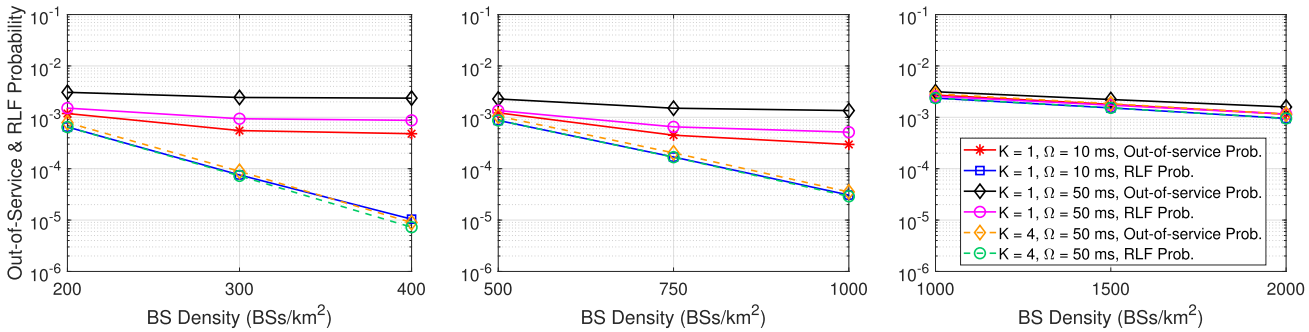
B. Out-of-Service Duration

Fig. 8 plots the out-of-service duration for different values of coverage range, BS density, connectivity degree, BS discovery time, and handover execution time for both mmWave and THz networks. We observe from Fig. 8 that the out-of-service duration can be significantly higher in the THz network as compared to the mmWave network, due to the following:

1) The Smaller Coverage Range and in turn the Smaller Number of BSs in the UE Coverage Region: We achieve lower average physical blockage duration with higher number of BSs in the UE coverage region. To understand that, let us consider a scenario with only one BS in the UE coverage region. In case of a blockage, the UE will be out-of-service unless and until the blockage ends and this single BS is discovered by the UE, which takes at least the sum of blockage duration and BS discovery time. On the contrary, if there are multiple BSs in the UE coverage region, the UE will be out-of-service until one BS gets discovered. The BS discovery by the UE after going out-of-service will require the minimum time before at least one BS gets unblocked and discovered, i.e., the earliest finishing time among all blockages and associated BS discovery.

2) Higher BS Discovery Time: Due to higher cell search time when a UE goes to the $[0, 0]$ state, the time to discover an unblocked BS will be much higher for a UE. Furthermore, with higher discovery time, it is more likely that another blocker will block the link before a new BS is discovered.

On the other hand, from Fig. 8, we observe that the expected out-of-service duration is larger for a higher multi-connectivity degree, which at first glance is counter-intuitive. This happens because, for a higher connectivity degree, the UE is out-of-service mostly when it is physically blocked, not protocol blocked as in the case of $K = 1$. For example, if we fix the BS density, the out-of-service probability reduces with the multi-connectivity degree (see Fig. 9) while the physical blockage probability (see Proposition 1) remains the same. The probability of blockage-by-protocol is higher when the multi-connectivity degree is lower (see Fig. 9). Therefore, many more instances of blockage-by-protocol will occur for a lower multi-connectivity degree. Note that in the mmWave and THz cellular networks, physical blockage duration can be quite large (typically a few hundreds of milliseconds [51]) as compared to blockage by protocol (typically a few tens of milliseconds [22]). Thus, from Fig. 8(a), Fig. 8(b), Fig. 9(a), and Fig. 9(b), we observe a lower average blockage duration for $K = 1$ than a higher connectivity degree $K > 1$, since in



(a) mmWave: $R = 100$ m. (b) THz: $R = 56$ m. (c) THz: $R = 23$ m.

Fig. 10. A comparison of the out-of-service probability and the RLF probability (Theorem 2). An illustration of the impact of the handover execution time Ω (10 and 50 ms) and multi-connectivity degree K (1 and 4) on RLF and out-of-service probability for BS density λ_T, blocker density λ_B (0.01 bl/m^2), BS discovery time Δ (20 ms).

a single connectivity ($K = 1$) scenario the average blockage duration is dominated by protocol blockages of short duration.

C. RLF Probability

Fig. 10 shows a comparison of the RLF probability and the out-of-service probability for different values of coverage range, BS density, connectivity degree, BS discovery and handover execution time.

From Fig. 10, we can observe that if the UE coverage range is large enough to have a sufficient number of BSs in its coverage region, reducing the handover execution time can significantly reduce the RLF probability (see Fig. 10(a) and Fig. 10(b)). However, if the UE has a small number of BSs in its coverage region, either due to smaller coverage range (see Fig. 10(c)) or lower BS density in the UE coverage region (see Fig. 10(a) and Fig. 10(b)), reducing the handover execution time does not help in reducing the RLF probability (see Fig. 10(c).) From Fig. 10(a) and Fig. 10(b), we further observe that having a small handover execution time, Ω, can significantly reduce the RLF probability in the single connectivity case and reaches close to the theoretical lower bound obtained in Corollary 7. Note that for $K = 1$ reducing the handover execution time is important, since in this scenario the UE will experience significantly more blockage-by-protocol events, which have an average duration of approximately Ω. Thus, a lower RLF probability can be achieved by reducing the handover execution time, Ω, below the RLF timer, $\mathcal{T}$, to avoid RLF declaration. From Fig. 10(a) and Fig. 10(b), we can further observe that even if the handover execution time is higher, for example $\Omega = 50$ ms, a higher degree connectivity results in similar performance to smaller handover execution time. *Therefore, a lower RLF probability can be obtained by either reducing the handover execution time or increasing the connectivity degree, provided that the UE has a sufficient number of BSs in its coverage region.*

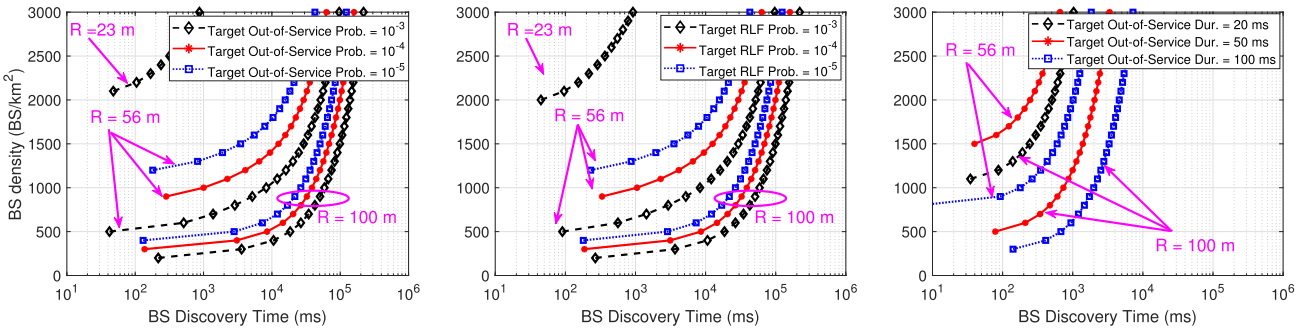
Recall from the above discussion and our discussion in V-B that in a scenario with a small number of BSs in the UE coverage region, the physical blockage probability is higher as compared to a scenario with a higher number of BSs (see Fig. 9). Moreover, the physical blockage instances have a much higher expected duration (a few hundred of milliseconds) than the blockage-by-protocol instances in the aforementioned scenario. Thus, in this scenario, the UE will almost always stay out-of-service for a duration longer than $\mathcal{T} = 50$ ms (see also Fig. 8) resulting in an RLF declaration.

D. Trade-off Analysis

Fig. 11 illustrates the minimum required BS density to achieve different target values of out-of-service and RLF probabilities, and out-of-service duration for a given coverage range and BS discovery time. Fig. 11 also illustrates the trade-off between coverage range and BS discovery time in THz cellular networks. In order to meet different QoS targets, the required number of BSs can be significantly reduced if a higher coverage range can be achieved in THz networks. However, as discussed earlier, compensating for the path loss requires a high beam directivity and consequently, the UE may have to spend a significantly higher amount of time for initial access and BS discovery procedures. For example, consider two THz cellular systems wherein one UE can communicate up to a range of 23 m whereas in another UE can communicate up to a range of 100 m by overcoming the path loss in THz systems. We observe from Fig. 11, the same target out-of-service and RLF requirement of 10^{-3} can be met with approximately one twelfth of the BS density (200 BSs/km^2 to 2400 BSs/km^2), in the long range THz system. However, as discussed earlier, in THz systems, it is not feasible to achieve a high coverage range with the same BS discovery budget. Nevertheless, even if the cost of using highly directed beams is a one to two orders of magnitude increase in the BS discovery time (e.g., 200 ms to 4000 ms), the same target out-of-service probability can still be achieved with one eighth of BSs density. Our initial results suggest that, *in THz cellular systems, the target QoS requirements can be achieved by a significantly lower BS density if the coverage range of the system can be increased, even when this increase comes at the expense of much higher initial access and BS discovery times.*

VI. Discussion

In the previous sections, we presented an analytical model to study the overall QoS performance in a blockage driven network and presented a detailed discussion for selected parameters in both mmWave and THz cellular networks.



(a) Trade-off for target out-of-service probability. (b) Trade-off for target RLF probability. (c) Trade-off for target out-of-service duration.

Fig. 11. Trade-off between the coverage range and BS discovery time for different target values of out-of-service probability, RLF probability, and out-of-service duration. To conduct the trade-off analysis for target out-of-service and RLF probabilities, we consider that the handover execution time (20 ms) is significantly smaller than the sum of blockage duration (500 ms) and BS discovery time (for any BS discovery time). For the out-of-service duration trade-off analysis, the handover execution time is considered as 20 ms. For all of the three scenarios, dynamic blocker density is 0.01 blockers/m^2 and RLF timer is $\mathcal{T} = 50$ ms.

There are three main QoS performance metrics, namely out-of-service probability, RLF probability, and out-of-service duration. Based upon application requirements, a network designer needs to consider the trade-off between these metrics. Furthermore, these QoS metrics are functions of UE coverage range, BS density, multi-connectivity degree, and BS discovery and handover execution times. While it is challenging to meet the QoS requirements in mmWave networks, it will be significantly more difficult to do so in THz networks due to their inherently short range and significantly higher BS discovery time. Our initial results suggest that, to reduce the required BS density in THz cellular systems, the coverage range should be increased, even when this increase comes at the cost of much higher initial access and BS discovery times.

Our discussion on the out-of-service and RLF probabilities suggests that dual connectivity is sufficient to meet the reliability requirements of emerging ultra-low latency and high reliability applications given a sufficiently high BS density, large UE coverage range, and small BS discovery and handover execution time. While meeting the reliability requirements poses significant challenges on the network, satisfying the latency requirement is even more challenging due to long blockage duration and difficulty in achieving the required cell search time. Thus, from a network designer prospective, if an application is tolerant to frequent short outages, then the network designer should consider reducing the BS discovery and handover execution times, instead of increasing the multi-connectivity degree. For example, to support applications like online gaming, voice-over-IP, and teleconferencing, where the average delay requirement is of the order of 50 ms or higher, single connectivity may be sufficient. Thus, a service provider can provision single connectivity and still meet the required QoS. In contrast, if the application can tolerate a few longer outages but is susceptible to frequent short outages, the network designer should consider increasing the multi-connectivity degree. For example, to support AR/VR applications, where frequent interruptions cannot be tolerated, a higher multi-connectivity degree may be desired. Note that the duration and frequency of physical blockage is driven by BS station density and UE coverage range. Finally, for applications that cannot tolerate frequent short out-of-service instances or infrequent long out-of-service instances, alternative techniques, such as time-sensitive networking, should be explored [52].

As the design of THz networks is still in an early study phase, many of the specific procedures and associated parameters are yet to be fully determined. The observations made in this work are based on the parameter selections presented in Table III. The presented modeling for protocol design can still be used by a cellular provider or UE vendor for a trade-off and system performance analysis even if the parameter values turn out to be different in future networks.

VII. Conclusion

THz wireless networks are expected to meet the QoS requirements of next-generation applications. Even though these higher carrier frequencies can provide high data rates, mmWave and THz links are quite susceptible to blockages, which will increase the need for frequent and fast handovers, as well as a higher degree of multi-connectivity. To study the impact of the handover latency and the multi-connectivity degree on service interruption and reliability, we designed a system model using a Markov chain that considers dynamic blockages, self-blockages, multi-connectivity, and BS discovery and handover execution times. The theoretical framework presented in this work can be used to evaluate the performance of any blockage driven network. The power of our theoretical modeling lies in the availability of analytic or closed form expressions allowing a network designer to quickly evaluate various trade-offs between coverage range, multi-connectivity degree, blocker density, BS deployment density, handover execution and BS discovery times. The results of this analysis are presented along with suggestions for future THz systems that stem from it.

Appendix

A. Proof of Proposition 1

Let us denote the number of unblocked BS-UE links by the binomial random variable, $L \sim \text{Binom}(M, q)$, where $q = \frac{\psi}{\alpha+\psi}$ and it represents a BS being unblocked and discovered by the UE in the 1×1 connectivity model. Let us also denote the

number of associated BSs as I.

$$\begin{aligned}\sum_{i=0}^{\min(l,K)} P_{li}^{M\times K} &= \sum_{i=0}^{\min(l,K)} P[L=l, I=i] \\ &= \sum_{i=0}^{\min(l,K)} P[L=l]P\left[I=i|L=l\right] \\ &= P[L=l]\sum_{i\in S} P\left[I=i|L=l\right] \\ &= \binom{M}{l}\left(\frac{\psi}{\alpha+\psi}\right)^{l}\left(\frac{\alpha}{\alpha+\psi}\right)^{M-l},\end{aligned}$$

where S is the set containing all possible realizations of I, thus the summation adds up to 1.

This concludes proof of Proposition 1

B. Proof of Proposition 2

Let us consider two distinct cases for Proposition 2:

1) $M < K$: If the number of BSs in the UE coverage region is less than maximum degree of connectivity UE can support, then the maximum achievable degree of connectivity, $\tilde{K}$, will be $\tilde{K} = M$.
2) $K = M$: In this case, the UE is allowed to achieve a maximum degree of connectivity equal to the number of BSs in the UE coverage region, i.e., UE can simultaneously connect to all of the BSs irrespectively of the connection status of the other BSs.

In fact, in both of the cases discussed above, the connectivity processes between the UE and the BSs can be considered as independent and identically distributed. Thus, the probability of state $[x, y]$ can be derived using (11) as

$$P_{xy}^{M\times K} = \binom{M}{x}\left[P_{00}^{1\times 1}\right]^{M-x}\binom{x}{y}\left[P_{10}^{1\times 1}\right]^{x-y}\left[P_{11}^{1\times 1}\right]^{y},$$

Now, the probability of the UE being associated to y BSs can be derived as

$$\begin{aligned}P_{\mathcal{C}y}^{M\times K} &= \sum_{x=y}^{M} P_{xy}^{M\times K} \\ &= \sum_{x=y}^{M}\binom{M}{x}\left[P_{00}^{1\times 1}\right]^{M-x}\binom{x}{y}\left[P_{10}^{1\times 1}\right]^{x-y}\left[P_{11}^{1\times 1}\right]^{y} \\ &= \left[P_{00}^{1\times 1}\right]^{M}\left[\frac{P_{11}^{1\times 1}}{P_{00}^{1\times 1}}\right]^{y}\sum_{k=0}^{M-y}\binom{M}{k+y}\binom{k+y}{y}\left[\frac{P_{10}^{1\times 1}}{P_{00}^{1\times 1}}\right]^{k} \\ &= \binom{M}{y}\left[P_{11}^{1\times 1}\right]^{y}\left[P_{00}^{1\times 1}+P_{10}^{1\times 1}\right]^{M-y} \\ &= \left[P_{00}^{1\times 1}+P_{10}^{1\times 1}\right]^{M}\binom{M}{y}\left[\frac{P_{11}^{1\times 1}}{P_{00}^{1\times 1}+P_{10}^{1\times 1}}\right]^{y} \quad (36)\end{aligned}$$

Finally, using (11) and (36), we obtain the probability of the UE being associated to y BSs as

$$P_{\mathcal{C}y}^{M\times K} = \left[\frac{\alpha}{\alpha+\omega}+\frac{\alpha}{\alpha+\psi}\frac{\omega}{\alpha+\omega}\right]^{M}\binom{M}{y}\left[\frac{\psi\omega}{\alpha(\alpha+\psi+\omega)}\right]^{y}$$

This concludes the proof of Proposition 2.

C. Proof of Proposition 3

For this proof, we will refer to all probabilities $P_{ij}^{M\times K}$ as P_{ij} for simplicity and readability. From the Markov chain in Fig. 4, we have $\forall i \in \{2, 3, \cdots, M-1\}$:

$$\begin{aligned}&\begin{bmatrix}-\alpha i\\ \min(K,i)w+(M-i)\psi+i\alpha\\ -(M-i+1)\psi\end{bmatrix}^{\top}\begin{bmatrix}P_{(i+1)0}\\ P_{i0}\\ P_{(i-1)0}\end{bmatrix} \\ &= \begin{bmatrix}-\alpha i\\ (M-i)\psi+i\alpha\\ -(M-i+1)\psi\end{bmatrix}^{\top}\begin{bmatrix}P_{i+1}\\ P_{i}\\ P_{i-1}\end{bmatrix} - \alpha\sum_{j=2}^{\min(i+1,K)} P_{(i+1)j} \quad (37)\end{aligned}$$

For the edge cases of the Markov chain, we can further write:

$$\begin{bmatrix}K\omega+M\alpha\\ -\psi\end{bmatrix}^{\top}\begin{bmatrix}P_{M0}\\ P_{(M-1)0}\end{bmatrix} = \begin{bmatrix}M\alpha\\ -\psi\end{bmatrix}^{\top}\begin{bmatrix}P_{M}\\ P_{M-1}\end{bmatrix} \quad (38)$$

and

$$\begin{aligned}&\begin{bmatrix}-\alpha\\ \omega+\alpha+(M-1)\psi\\ \omega+\alpha\end{bmatrix}^{\top}\begin{bmatrix}P_{20}\\ P_{10}\\ P_{00}\end{bmatrix} = \begin{bmatrix}-\alpha\\ (M-1)\psi+\alpha\\ \omega+\alpha\end{bmatrix}^{\top} \\ &\qquad\qquad \times \begin{bmatrix}P_{2}\\ P_{1}\\ P_{0}\end{bmatrix} - \alpha P_{22} \quad (39)\end{aligned}$$

By adding equations (37), (38), and (39) together, we have:

$$\begin{aligned}&(\omega+\alpha)(P_{M0}+\cdots+P_{00}) \\ &= \alpha(P_M+\cdots+P_0)+\omega P_0 \\ &\quad -\alpha\sum_{a=2}^{M}\sum_{b=2}^{min(a+1,K)} P_{ab} - (K-1)\omega\sum_{a=K}^{M} P_{a0} - \omega\sum_{a=2}^{K-1} aP_{a0} \\ &\Rightarrow (\omega+\alpha)P_{OS} = \alpha+\omega P_0 - \alpha\sum_{a=2}^{M}\sum_{b=2}^{\min(a+1,K)} P_{ab} \\ &\quad -\sum_{a=2}^{M}\min(a-1,K-1)P_{a0} \\ &\Rightarrow P_{OS} = P_0 - \sum_{a=2}^{M}\min(a-1,K-1)P_{a0} + \frac{\alpha}{\omega}P_{C1}.\end{aligned}$$

This concludes the proof of Proposition 3.

D. Proof of Corollary 3

In Proposition 3, we derived the relationship among out-of-service states and other states. Finding the exact expression of out-of-service probability in the $M \times K$ connectivity setting was challenging. Instead, we derived upper and lower bounds on the out-of-service probability in Corollary 3. From Proposition 3, the relation is derived as:

$$P_{OS}^{M\times K} = P_0^{M\times K} - \sum_{i=2}^{M}\min(i-1,K-1)P_{i0}^{M\times K} + \frac{\alpha}{\omega}P_{C1}^{M\times K}.$$

In the above equation, to find the upper bound, we replaced $\min(i-1, K-1)$ with 1, as it takes the minimum value of 1. Furthermore, $\sum_{i=2}^{M}\min(i-1,K-1)P_{i0}^{M\times K} \geq \sum_{i=2}^{M} P_{i0}^{M\times K} = P_{OS}^{M\times K} - \left(P_0^{M\times K}+P_{10}^{M\times K}\right)$. Thus, we get

$$P_{OS}^{M\times K} \leq P_0^{M\times K} + \frac{1}{2}P_{10}^{M\times K} + \frac{\alpha}{2\omega}P_{C1}^{M\times K}.$$

Now, we argue that if M is significantly larger than K, then $P_{10}^{M\times K} \leq P_{10}^{K\times K}$ and $P_{C1}^{M\times K} \leq P_{C1}^{K\times K}$. This happens because in the $M \times K$ scenario a large number of new states are introduced as compared to the $K \times K$ scenario, thus the rate of transitioning to the old states will become smaller. Therefore, we obtain the upper bound as:

$$P_{OS,UB}^{M\times K} \leq \left[\frac{\alpha}{\alpha+\psi}\right]^M + \frac{K}{2}\left[\frac{\alpha}{\alpha+\psi}\right]^K \left[\frac{\psi}{\alpha+\omega}\right] + \frac{K\alpha}{2\omega}\left[\frac{\alpha}{\alpha+\omega} + \frac{\alpha}{\alpha+\psi}\frac{\omega}{\alpha+\omega}\right]^{K-1}\left[\frac{\psi\omega}{(\alpha+\psi)(\alpha+\omega)}\right].$$

To find the lower bound, we replaced $\min(i-1, K-1)$ with $K-1$, as it takes maximum value as $K-1$. Furthermore, $\sum_{i=2}^{M} \min(i-1,K-1)P_{i0}^{M\times K} \leq \sum_{i=2}^{M}(K-1)P_{i0}^{M\times K}$ and $\sum_{i=2}^{M} P_{i0}^{M\times K} = P_{OS}^{M\times K} - \left(P_0^{M\times K} + P_{10}^{M\times K}\right)$. Thus, we get

$$P_{OS}^{M\times K} \geq P_0^{M\times K} + \frac{K-1}{K}P_{10}^{M\times K} + \frac{\alpha}{K\omega}P_{C1}^{M\times K}.$$

Now, we argue that if M is significantly larger than K, then $P_{10}^{M\times K} \geq P_{10}^{M\times M}$ and $P_{C1}^{M\times K} \geq P_{C1}^{M\times M}$. This happens because after removing a large number of states from $M \times M$ connectivity setting, we get the $M \times K$ connectivity setting, thus the state probabilities in the $M \times K$ connectivity setting will be higher than $M \times M$ for particular states. Thus, we get the lower bound as:

$$P_{OS,LB}^{M\times K} \geq \left[\frac{\alpha}{\alpha+\psi}\right]^M + M\frac{K-1}{K}\left[\frac{\alpha}{\alpha+\psi}\right]^M \left[\frac{\psi}{\alpha+\omega}\right] + \frac{M\alpha}{K\omega}\left[\frac{\alpha}{\alpha+\omega} + \frac{\alpha}{\alpha+\psi}\frac{\omega}{\alpha+\omega}\right]^{M-1}\left[\frac{\psi\omega}{(\alpha+\psi)(\alpha+\omega)}\right].$$

This concludes the proof of Corollary 3.

E. Proof of Theorem 1

Let us first derive the marginal out-of-service probability for $K = 1$. Using (5) and (20), we can derive the marginal out-of-service probability as

$$\begin{aligned} P(OS) &= \sum_{m=1}^{\infty} P(OS|m)P_M(m) \\ &= \sum_{m=1}^{\infty}\left[\frac{\alpha}{\alpha+\omega} + \frac{\omega}{\alpha+\omega}\left(\frac{\alpha}{\alpha+\psi}\right)^m\right] \times \frac{\left[p\lambda_T\pi R^2\right]^m}{m!}e^{-p\lambda_T\pi R^2} \\ &= \frac{\alpha}{\alpha+\omega}\sum_{m=1}^{\infty}\frac{\left[p\lambda_T\pi R^2\right]^m}{m!}e^{-p\lambda_T\pi R^2} + \frac{\omega}{\alpha+\omega} \times \sum_{m=1}^{\infty}\left(\frac{\alpha}{\alpha+\psi}\right)^m \times \frac{\left[p\lambda_T\pi R^2\right]^m}{m!}e^{-p\lambda_T\pi R^2} \\ &= \frac{\alpha}{\alpha+\omega}\left(1-e^{-p\lambda_T\pi R^2}\right) + \frac{\omega}{\alpha+\omega}\left(e^{-qp\lambda_T\pi R^2} - e^{-p\lambda_T\pi R^2}\right) \\ &= 1 - \frac{\omega}{\omega+\alpha}\left(1-e^{-qp\lambda_T\pi R^2}\right) - e^{-p\lambda_T\pi R^2}. \end{aligned} \quad (40)$$

Furthermore, the out-of-service probability given LOS coverage is derive using Bayes's rule, (6), and (40) as

$$\begin{aligned} P(OS) &= P(OS|C^{\text{LOS}})P(C^{\text{LOS}}) \Rightarrow P(OS|C^{\text{LOS}}) \\ &= \frac{P(OS)}{P(C^{\text{LOS}})}. \end{aligned} \quad (41)$$

Thus,

$$P\left(OS|C^{\text{LOS}}\right) = \frac{1 - \frac{\omega}{\alpha+\omega}\left(1-e^{-qp\lambda_T\pi R^2}\right) - e^{-p\lambda_T\pi R^2}}{1-e^{-p\lambda_T\pi R^2}}, \quad \text{if } K = 1,$$

where $q = \frac{\psi}{\alpha+\psi}$. Furthermore, using (20), (41), and similar steps as in $K = 1$, we can prove upper and lower bounds on the out-of-service probability given LOS coverage for $K \geq 2$. This concludes the proof of Theorem 1.

F. Proof of Corollary 4

If $\omega \gg \alpha$, $\omega \gg \psi$, then $\frac{\psi}{\alpha+\omega} \to 0 \Rightarrow \chi \to 0$, $\frac{\alpha}{\alpha+\omega} \to 0 \Rightarrow P_{0S}^{1\times 1} \to \frac{\alpha}{\alpha+\mu}, and \frac{m\alpha\zeta}{K\omega} \to 0$. Thus, we get

$$\begin{aligned} P\left(OS|C^{\text{LOS}}\right) &= \frac{e^{-qp\lambda_T\pi R^2} - e^{-p\lambda_T\pi R^2}}{1-e^{-p\lambda_T\pi R^2}}; \text{ if } K = 1 \\ &\leq \frac{e^{-p\lambda_T\pi R^2}}{1-e^{-p\lambda_T\pi R^2}}\left[\sum_{m=1}^{\infty}\frac{\left[\tilde{q}p\lambda_T\pi R^2\right]^m}{m!}\right]; \\ &\quad \text{if } K \geq 2. \\ &\geq \frac{e^{-p\lambda_T\pi R^2}}{1-e^{-p\lambda_T\pi R^2}}\left[\sum_{m=1}^{\infty}\frac{\left[\tilde{q}p\lambda_T\pi R^2\right]^m}{m!}\right]; \\ &\quad \text{if } K \geq 2. \end{aligned}$$

Therefore, for small handover execution times, we obtain the out-of-service probability given LOS coverage as

$$P\left(OS|C^{\text{LOS}}\right) = \frac{e^{-qp\lambda_T\pi R^2} - e^{-p\lambda_T\pi R^2}}{1-e^{-p\lambda_T\pi R^2}},$$

irrespective of the multi-connectivity degree. This concludes the proof of Corollary 4.

G. Proof of Corollary 5

Using Corollary 4, we have

$$\begin{aligned} &e^{-qp\lambda_T\pi R^2} = P\left(OS|C^{\text{LOS}}\right) + P\left(OS|C^{\text{LOS}}\right)\left(1-P\left(OS|C^{\text{LOS}}\right)\right) \\ &\Rightarrow q = -\frac{\ln\left[P\left(OS|C^{\text{LOS}}\right)+P\left(OS|C^{\text{LOS}}\right)\left(1-P\left(OS|C^{\text{LOS}}\right)\right)\right]}{p\lambda_T\pi R^2} \\ &\Rightarrow \frac{\psi}{\alpha+\psi} = -\Upsilon \Rightarrow \Psi = -\frac{1+\Upsilon}{\Upsilon\alpha} \Rightarrow \Delta = -\frac{1+\Upsilon}{\Upsilon\alpha} - \Gamma, \\ &\text{where } \Upsilon = \frac{ln\left[P\left(OS|C^{\text{LOS}}\right) + e^{-p\lambda_T\pi R^2}\left(1-P\left(OS|C^{\text{LOS}}\right)\right)\right]}{p\lambda_T\pi R^2}. \end{aligned}$$

This concludes the proof.

H. Proof of Corollary 6

From (21), we have

$$\begin{aligned}&\lim_{K\to\infty} P\left(OS|C^{\text{LOS}}\right)\\&\leq \lim_{K\to\infty}\left[\frac{e^{-p\lambda_T\pi R^2}}{1-e^{-p\lambda_T\pi R^2}}\sum_{m=1}^{K}\frac{\left[cp\lambda_T\pi R^2\right]^m}{m!}\right]\\&\geq \lim_{K\to\infty}\left[\frac{e^{-p\lambda_T\pi R^2}}{1-e^{-p\lambda_T\pi R^2}}\sum_{m=1}^{K}\frac{\left[cp\lambda_T\pi R^2\right]^m}{m!}\right]\end{aligned}\tag{42}$$

Thus, from (42), we have

$$\begin{aligned}&\lim_{K\to\infty} P\left(OS|C^{\text{LOS}}\right)\\&=\lim_{K\to\infty}\left[\frac{e^{-p\lambda_T\pi R^2}}{1-e^{-p\lambda_T\pi R^2}}\sum_{m=1}^{K}\frac{\left[cp\lambda_T\pi R^2\right]^m}{m!}\right]\\&=\frac{e^{-p\lambda_T\pi R^2}}{1-e^{-p\lambda_T\pi R^2}}\left(e^{-cp\lambda_T\pi R^2}-1\right)\\&=\frac{e^{-\tilde{c}p\lambda_T\pi R^2}-e^{-p\lambda_T\pi R^2}}{1-e^{-p\lambda_T\pi R^2}},\end{aligned}$$

where $\tilde{c}=1-c$. This concludes the proof.

I. Proof of Theorem 2

Using (5), (20), and (25), we have

$$\begin{aligned}&P\left(RLF\right)\\&\geq\sum_{m=1}^{\infty}P_{0S}^{M\times K}e^{-K\omega\mathcal{T}}\frac{\left[p\lambda_T\pi R^2\right]^m}{m!}e^{-p\lambda_T\pi R^2}\\&\quad+\sum_{m=1}^{\infty}P_{00}^{M\times K}\left(e^{-m\psi\mathcal{T}}-e^{-K\omega\mathcal{T}}\right)\frac{\left[p\lambda_T\pi R^2\right]^m}{m!}e^{-p\lambda_T\pi R^2}\\&=P(OS)e^{-K\omega\mathcal{T}}+\sum_{m=1}^{\infty}(\tilde{q})^m\left(e^{-m\psi\mathcal{T}}-e^{-K\omega\mathcal{T}}\right)\\&\quad\times\frac{\left[p\lambda_T\pi R^2\right]^m}{m!}e^{-p\lambda_T\pi R^2}\\&=P(OS)e^{-K\omega\mathcal{T}}+\left[\sum_{m=1}^{\infty}\frac{\left[\tilde{q}e^{-\psi\mathcal{T}}p\lambda_T\pi R^2\right]^m}{m!}-e^{-K\omega T}\right.\\&\quad\left.\times\sum_{m=1}^{\infty}\frac{\left[\tilde{q}p\lambda_T\pi R^2\right]^m}{m!}\right]e^{-p\lambda_T\pi R^2}=P(OS)e^{-K\omega\mathcal{T}}\\&\quad+\left(e^{\xi\tilde{q}p\lambda_T\pi R^2}-e^{-K\omega T}\left(e^{\tilde{q}p\lambda_T\pi R^2}-1\right)-1\right)\\&\quad\times e^{-p\lambda_T\pi R^2}\end{aligned}\tag{43}$$

Furthermore, using (41) and (43), we have

$$\begin{aligned}P\left(RLF|C^{\text{LOS}}\right)\geq{}&P\left(OS|C^{\text{LOS}}\right)e^{-K\omega\mathcal{T}}\\&+\frac{e^{-p\lambda_T\pi R^2}}{1-e^{-p\lambda_T\pi R^2}}\left[e^{\xi\tilde{q}p\lambda_T\pi R^2}-\right.\\&\left.e^{-K\omega T}\left(e^{\tilde{q}p\lambda_T\pi R^2}-1\right)-1\right],\end{aligned}$$

where $\xi=e^{-\psi\mathcal{T}}$ and $\tilde{q}=1-q=\frac{\alpha}{\alpha+\psi}$. This concludes the proof of Theorem 2.

J. Proof of Corollary 7

From Corollary 4 and Theorem 2, we have

$$\begin{aligned}&P\left(RLF|C^{\text{LOS}}\right)\\&\geq\frac{e^{-p\lambda_T\pi R^2}}{1-e^{-p\lambda_T\pi R^2}}\left(e^{\tilde{q}p\lambda_T pi R^2}-1\right)\\&\quad\times e^{-K\omega\mathcal{T}}+\frac{e^{-p\lambda_T\pi R^2}}{1-e^{-p\lambda_T\pi R^2}}\\&\quad\times\left[e^{\xi\tilde{q}p\lambda_T\pi R^2}-e^{-K\omega T}\left(e^{\tilde{q}p\lambda_T\pi R^2}-1\right)-1\right]\\&=\frac{e^{-p\lambda_T\pi R^2}}{1-e^{-p\lambda_T\pi R^2}}\left[e^{\xi\tilde{q}p\lambda_T\pi R^2}-1\right]\\&=\frac{e^{-(1-\xi\tilde{q})p\lambda_T\pi R^2}-e^{-p\lambda_T\pi R^2}}{1-e^{-p\lambda_T\pi R^2}}.\end{aligned}$$

This concludes the proof of Corollary 7.

K. Proof of Corollary 8

The proof follows similar steps as presented in proof of Corollary 5 and we get

$$\frac{\alpha e^{-\psi\mathcal{T}}}{\alpha}+\psi=1+\eta$$

where,

$$\eta=\frac{ln\left[P\left(RLF|C^{\text{LOS}}\right)+e^{-p\lambda_T\pi R^2}\left(1-P\left(RLF|C^{\text{LOS}}\right)\right)\right]}{p\lambda_T\pi R^2}\tag{44}$$

By solving (44), we get

$$\Delta=\frac{\alpha}{W_n\left(\frac{\alpha\mathcal{T}e^{\alpha\mathcal{T}}}{\eta}\right)-\alpha\mathcal{T}}-\Gamma,$$

where $W_k(z)$ is the analytic computation of the product log function. This concludes the proof.

L. Proof of Corollary 9

From Corollary 7, we have

$$P\left(RLF|C^{\text{LOS}}\right)\geq\frac{e^{-p\lambda_T\pi R^2}}{1-e^{-p\lambda_T\pi R^2}}\left[e^{\xi\tilde{q}p\lambda_T\pi R^2}-1\right]\tag{45}$$

Now, using Corollary 4 and (45), we have

$$\frac{P\left(RLF|C^{\text{LOS}}\right)}{P\left(OS|C^{\text{LOS}}\right)}\geq\frac{e^{\xi\tilde{q}p\lambda_T\pi R^2}-1}{e^{\tilde{q}p\lambda_T\pi R^2}-1}\approx\left(e^{-\tilde{q}p\lambda_T\pi R^2}\right)^{1-\xi},$$

where the approximation is obtained by considering $e^{\xi\tilde{q}p\lambda_T\pi R^2}\gg 1$ and $e^{\tilde{q}p\lambda_T\pi R^2}\gg 1$. This concludes the proof of Corollary 9.

M. Details of the Out-of-Service Duration $T_{OS}^{M\times K}$ Computation

Recall from Fig. 4 that if the UE is in service, it can only get to an out-of-service state [i,0] via the connected state [i+1,1]. Let us denote by A_i the event of the UE going to the out-of-service state $[i,0]$ via the state $[i+1,1], i=0,\cdots,M-1$.

Then, we can write the event of the UE going to an out-of-service state while being in service as:

$$A = \bigcup_{i=0}^{M-1} A_i$$

To obtain the expected out-of-service duration $T_{OS}^{M\times K}$, we need to calculate the expected time that the UE spent in out-of-service states starting from the instant the UE entered an out-of-service state via an in-service state, i.e., by conditioning on the event A.

$$\begin{aligned} T_{OS}^{M\times K} &= \mathbb{E}\left[t_{OS}^{M\times K}|A\right] \\ &= \sum_{k=0}^{M-1} \mathbb{E}\left[t_{OS}^{M\times K}|A, A_k\right] \times P(A_k|A) \\ &= \sum_{k=0}^{M-1} \mathbb{E}\left[t_{OS}^{M\times K}|A_k\right] \times \frac{P(A_k)}{P(A)} \\ &= \sum_{k=0}^{M-1} \mathbf{t}^{\mathbf{M}\times\mathbf{K}}([k,0]) \times \frac{P(A_k)}{\sum_{j=0}^{M-1} P(A_j)} \end{aligned} \quad (46)$$

However, the event $A_k,\ k = 0, \cdots, M-1$ is equivalent to the event of 1) being in the state $[K+1, 1]$, and 2) transitioning from $[K+1, 1]$ to $[k, 0]$. Thus,

$$\begin{aligned} P(A_k) &= P_{k+1,1} \times \frac{\alpha}{(\eta_k - 1)\omega + (M-k-1)\psi + k\alpha} \\ &= P_{k+1,1} \times \alpha\, \mathbf{t}_{\text{leave}}^{\mathbf{M}\times\mathbf{K}}([k,1]) \end{aligned} \quad (47)$$

By substituting $P(A_k)$ in (46) with (47), we obtain the expected out-of-service duration.

N. Proof of Lemma 2

To obtain the expected out-of-service duration using Little's Law, we can envision our $M \times K$ Markov chain as a simplified queueing system. In this queueing system, there are two states, *connected* and *blocked*. The customer in this system arrives (enters the blocked state) with rate

$$r_b = \alpha P_{C1}^{M\times K}.$$

Once a customer arrives, the system goes to outage. There cannot be new customers until the existing customer gets served, i.e., the system gets back to a connected state. The average number of customers in the system is

$$E_b = 1 \times P_{OS}^{M\times K} = P_{OS}^{M\times K}$$

By applying the Little's Law into the aforementioned system we get

$$T_{OS}^{M\times K} = \frac{E_b}{r_b} = \frac{P_{OS}^{M\times K}}{\alpha P_{C1}^{M\times K}}.$$

This concludes the proof of Lemma 2.

O. Proof of Theorem 3

From Lemma 2, we have

$$\mathbb{E}\left[T_{OS}^{M\times K}\right] = \sum_{m=1}^{\infty} T_{OS}^{M\times K} P_M(m) = \sum_{m=1}^{\infty} \frac{P_{OS}^{m\times K}}{\alpha P_{C1}^{m\times K}} P_M(m) \quad (48)$$

Let us first derive the out-of-service duration for $K = 1$. Using (48), we have:

$$\begin{aligned} &\mathbb{E}\left[T_{OS}^{m\times 1}\right] \\ &= \sum_{m=1}^{\infty} T_{OS}^{m\times 1} P_M(m) \\ &= \sum_{m=1}^{\infty} \frac{\frac{\alpha}{\alpha+\omega} + (\frac{\alpha}{\alpha+\psi})^m \frac{\omega}{\alpha+\omega}}{\alpha\left[\frac{\omega}{\alpha+\omega} - (\frac{\alpha}{\alpha+\psi})^m \frac{\omega}{\alpha+\omega}\right]} \frac{[p\lambda_T\pi R^2]^m}{m!} e^{-p\lambda_T\pi R^2} \\ &= \sum_{m=1}^{\infty} \frac{\frac{\omega}{\alpha+\omega}\left[\frac{\alpha}{\omega} + \left(\frac{\alpha}{\alpha+\psi}\right)^m\right]}{\frac{\alpha\omega}{\alpha+\omega}\left[1 - \left(\frac{\alpha}{\alpha+\psi}\right)^m\right]} \frac{[p\lambda_T\pi R^2]^m}{m!} e^{-p\lambda_T\pi R^2} \\ &\simeq \left[\frac{\alpha}{\omega}\sum_{m=1}^{\infty} \frac{[p\lambda_T\pi R^2]^m}{m!} + \sum_{m=1}^{\infty} \frac{[\tilde{q}p\lambda_T\pi R^2]^m}{m!}\right] \frac{e^{-p\lambda_T\pi R^2}}{\alpha} \\ &= \left[\frac{\alpha}{\omega}[e^{p\lambda_T\pi R^2} - 1] + e^{\tilde{q}p\lambda_T\pi R^2} - 1\right] \frac{e^{-p\lambda_T\pi R^2}}{\alpha} \\ &= \frac{1}{\omega}\left[1 - e^{-p\lambda_T\pi R^2}\right] + \frac{1}{\alpha}\left[e^{-(1-\tilde{q})p\lambda_T\pi R^2} - e^{-p\lambda_T\pi R^2}\right] \end{aligned}$$

Now for $K \geq 2$, using Proposition 3, we can obtain the expected out of service duration for $M \times K$ connectivity model, $\mathbb{E}\left[T_{OS}^{m\times k}\right]$ in a similar fashion as described above. By further using the fact that

$$\mathbb{E}\left[T\left(OS|C^{\text{LOS}}\right)\right] = \frac{\mathbb{E}\left[T\left(OS\right)\right]}{P(C^{\text{LOS}})},$$

we obtain expression (34). This concludes Theorem 3.

P. Proof of Corollary 10

When $K \to \infty$, from Theorem 3, we have

$$\mathbb{E}\left[T\left(OS|C^{\text{LOS}}\right)\right] = \frac{e^{-p\lambda_T\pi R^2}}{1 - e^{-p\lambda_T\pi R^2}}\left[\frac{\alpha+\psi+\omega}{\psi\omega}\text{Ei}\left[p\lambda_T\pi R^2\right]\right]$$

$$\Rightarrow \psi = \frac{\alpha+\omega}{\nu\omega - 1} \Rightarrow \Delta = \frac{\nu\omega - 1}{\alpha+\omega} - \Gamma,$$

where $\nu = \frac{\mathbb{E}\left[T\left(OS|C^{\text{LOS}}\right)\right]\left(1 - e^{-p\lambda_T\pi R^2}\right)}{\text{Ei}[p\lambda_T\pi R^2]e^{-p\lambda_T\pi R^2}}$ and $\text{Ei}\left[p\lambda_T\pi R^2\right] = \sum_{m=1}^{\infty} \frac{[p\lambda_T\pi R^2]^m}{m m!}$.

This concludes the proof.

ACKNOWLEDGMENT

The author would like to thank the editors for coordinating the review process, and the anonymous reviewers for their insightful comments and suggestions.

REFERENCES

[1] NGM Alliance, "NGMN 5G initiative white paper," NGMN, Frankfurt, Germany, Tech. Rep., Feb. 2015. [Online]. Available: https://tinyurl.com/m6889bkt

[2] *Service Requirements for 5G System*, Standard 3GPP TS 22.261, 3GPP Std. v16.8.0, Jun. 2019.

[3] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work," *IEEE Access*, vol. 1, pp. 335–349, 2013.

[4] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," 2019, *arXiv:1902.10265*. [Online]. Available: http://arxiv.org/abs/1902.10265

[5] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5G wireless network slicing for eMBB, URLLC, and mMTC: A communication-theoretic view," *IEEE Access*, vol. 6, pp. 55765–55779, 2018.

[6] Y. Rao *et al.*, "New services & applications with 5G ultra-reliable low latency communication," 5G Americas, Bellevue, WA, USA, Tech. Rep., Nov. 2018.

[7] Cisco, "Cisco visual networking index: Forecast and trends, 2017–2022," Cisco, San Jose, CA, USA, Tech. Rep., Dec. 2018. [Online]. Available: https://tinyurl.com/hj7htmvb

[8] Future Networks Team, Huawei Technologies, USA, "Towards a new Internet for the year 2030 and beyond," Huawei, Shenzhen, China, Tech. Rep. ITU-T, SG 13, Jul. 2018. [Online]. Available: https://tinyurl.com/yxu7t66q

[9] Z. Zhang *et al.*, "6G wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 28–41, Sep. 2019.

[10] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, "6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions," 2019, *arXiv:1909.11315*. [Online]. Available: http://arxiv.org/abs/1909.11315

[11] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, Mar. 2020.

[12] T. S. Rappaport *et al.*, "Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond," *IEEE Access*, vol. 7, pp. 78729–78757, 2019.

[13] K. Allen, N. DeMinco, J. Hoffman, Y. Lo, and P. Papazian, "Building penetration loss measurements at 900 MHz, 11.4 GHz, and 28.8 GHz," U.S. Dept. Commerce, Nat. Telecommun. Inf. Admin., Boulder, CO, USA, Tech. Rep. 94-306, 1994.

[14] I. K. Jain, R. Kumar, and S. Panwar, "Driven by capacity or blockage? A millimeter wave blockage analysis," in *Proc. 30th Int. Teletraffic Congr. (ITC)*, Sep. 2018, pp. 153–159.

[15] I. K. Jain, R. Kumar, and S. S. Panwar, "The impact of mobile blockers on millimeter wave cellular systems," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 854–868, Apr. 2019.

[16] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: Opportunities and challenges," *Wireless Netw.*, vol. 21, no. 8, pp. 2657–2676, Nov. 2015.

[17] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[18] J. M. Jornet and I. F. Akyildiz, "Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3211–3221, Oct. 2011.

[19] V. Petrov, D. Moltchanov, Y. Koucheryavy, and J. M. Jornet, "The effect of small-scale mobility on terahertz band communications," in *Proc. 5th ACM Int. Conf. Nanosc. Comput. Commun.*, Sep. 2018, pp. 1–2.

[20] C. N. Barati *et al.*, "Initial access in millimeter wave cellular systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 7926–7940, Dec. 2016.

[21] C. N. Barati, S. Dutta, S. Rangan, and A. Sabharwal, "Energy and latency of beamforming architectures for initial access in mmWave wireless networks," *J. Indian Inst. Sci.*, vol. 100, no. 2, pp. 281–302, May 2020.

[22] H.-S. Park, Y. Lee, T.-J. Kim, B.-C. Kim, and J.-Y. Lee, "Handover mechanism in NR for ultra-reliable low-latency communications," *IEEE Netw.*, vol. 32, no. 2, pp. 41–47, Mar. 2018.

[23] H.-S. Park, Y.-S. Choi, B.-C. Kim, and J.-Y. Lee, "LTE mobility enhancements for evolution into 5G," *ETRI J.*, vol. 37, no. 6, pp. 1065–1076, Dec. 2015.

[24] *Requirements for Support of Radio Resource Management*, Standard 3GPP TS 36.133, 3GPP Std. v14.3.0, Apr. 2017.

[25] X. Yan, Y. A. Şekercioğlu, and S. Narayanan, "A survey of vertical handover decision algorithms in fourth generation heterogeneous wireless networks," *Comput. Netw.*, vol. 54, no. 11, pp. 1848–1863, Aug. 2010.

[26] *Overall Description*, Standard 3GPP TS 36.300, 3GPP Std. v12.5.0, Mar. 2015.

[27] "The 5G evolution: 3GPP Release 16–17," 5G Americas, Bellevue, WA, USA, Tech. Rep. Rel. 16-17, Jan. 2020. [Online]. Available: https://www.5gamericas.org/wp-content/uploads/2020/01/5G-Evolution-3GPP%-R16-R17-FINAL.pdf

[28] *Multi-Connectivity; Overall Description*, Standard 3GPP TS 37.340, 3GPP Std. v16.1.0, Apr. 2020.

[29] H. Wang, C. Rosa, and K. I. Pedersen, "Dual connectivity for LTE-advanced heterogeneous networks," *Wireless Netw.*, vol. 22, no. 4, pp. 1315–1328, May 2016.

[30] M. Gapeyenko *et al.*, "On the degree of multi-connectivity in 5G millimeter-wave cellular urban deployments," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1973–1978, Feb. 2019.

[31] A. Koutsaftis, R. Kumar, P. Liu, and S. S. Panwar, "Fast inter-base station ring (FIBR): A new millimeter wave cellular network architecture," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 12, pp. 2699–2714, Dec. 2019.

[32] M. Polese, M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "Improved handover through dual connectivity in 5G mmWave mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2069–2084, Sep. 2017.

[33] D. H. Hagos and R. Kapitza, "Study on performance-centric offload strategies for LTE networks," in *Proc. 6th Joint IFIP Wireless Mobile Netw. Conf. (WMNC)*, Apr. 2013, pp. 1–10.

[34] V. Petrov *et al.*, "Dynamic multi-connectivity performance in ultra-dense urban mmWave deployments," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2038–2055, Sep. 2017.

[35] B. Zhang, W. Qi, and J. Zhang, "A Markov based performance analysis of handover and load balancing in HetNets," *Int. J. Commun., Netw. Syst. Sci.*, vol. 10, no. 10, pp. 223–233, 2017.

[36] F. Guidolin, I. Pappalardo, A. Zanella, and M. Zorzi, "Context-aware handover policies in HetNets," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1895–1906, Mar. 2016.

[37] M. Gerasimenko, D. Moltchanov, M. Gapeyenko, S. Andreev, and Y. Koucheryavy, "Capacity of multiconnectivity mmWave systems with dynamic blockage and directional antennas," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3534–3549, Apr. 2019.

[38] A. Shafie, N. Yang, and C. Han, "Multi-connectivity for indoor terahertz communication with self and dynamic blockage," 2020, *arXiv:2004.07469*. [Online]. Available: http://arxiv.org/abs/2004.07469

[39] V. Begishev *et al.*, "Performance analysis of multi-band microwave and millimeter-wave operation in 5G NR systems," *IEEE Trans. Wireless Commun.*, early access, Jan. 20, 2021, doi: 10.1109/TWC.2021.3051027.

[40] V. Begishev *et al.*, "Quantifying the impact of guard capacity on session continuity in 3GPP new radio systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 12345–12359, Dec. 2019.

[41] M. Polese, J. Jornet, T. Melodia, and M. Zorzi, "Toward end-to-end, full-stack 6G terahertz networks," 2020, *arXiv:2005.07989*. [Online]. Available: http://arxiv.org/abs/2005.07989

[42] *Radio Resource Control (RRC); Protocol Specification*, Standard tandard3GPP TS 38.331, 3GPP Std. v16.0.0, Apr. 2020.

[43] A. Zimmermann and G. Hommel, "Towards modeling and evaluation of ETCS real-time communication and operation," *J. Syst. Softw.*, vol. 77, no. 1, pp. 47–54, Jul. 2005.

[44] K. T. P. Nguyen, J. Beugin, M. Berbineau, and M. Kassab, "A new analytical approach to evaluate the critical-event probability due to wireless communication errors in train control systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 6, pp. 1380–1392, Jun. 2017.

[45] LTE Quick Reference. *Radio Link Failure (RLF)*. Accessed: May 4, 2019. [Online]. Available: https://bit.ly/2lzjNLb

[46] S. M. Ross, *Introduction to Probability Models*. New York, NY, USA: Academic, 1997.

[47] J. D. Little and S. C. Graves, "Little's law," in *Building Intuition*. Boston, MA, USA: Springer, 2008, pp. 81–100.

[48] P. Skrimponis *et al.*, "Power consumption analysis for mobile mmWave and sub-THz receivers," in *Proc. 2nd 6G Wireless Summit (6G SUMMIT)*, Mar. 2020, pp. 1–5.

[49] D. B. Johnson and D. A. Maltz, *Dynamic Source Routing in Ad Hoc Wireless Networks*. Boston, MA, USA: Springer, 1996, pp. 153–181.

[50] M. Boutin. *Random Waypoint Mobility Model*. Accessed: Mar. 18, 2019. [Online]. Available: https://www.mathworks.com

[51] G. R. MacCartney, T. S. Rappaport, and S. Rangan, "Rapid fading due to human blockage in pedestrian crowds at 5G millimeter-wave frequencies," in *Proc. GLOBECOM IEEE Global Commun. Conf.*, Dec. 2017, pp. 1–7.

[52] N. Finn, "Introduction to time-sensitive networking," *IEEE Commun. Standards Mag.*, vol. 2, no. 2, pp. 22–28, Jun. 2018.

Mustafa F. Özkoç (Student Member, IEEE) received the B.Sc. degree in electrical and electronics engineering and physics from Bogazici University, Istanbul, Turkey, in 2018. He is currently pursuing the Ph.D. degree in electrical engineering with the NYU Tandon School of Engineering. He worked with Nokia Bell Labs during the summer of 2020. His research interests include wireless communications, wireless networking, and signal processing.

Athanasios Koutsaftis (Student Member, IEEE) received the B.Tech. degree in electrical and computer engineering from the National Technical University of Athens in 2015. He is currently pursuing the Ph.D. degree in electrical engineering with the Tandon School of Engineering, New York University, NY, USA. He was awarded the Dean's Fellowship to pursue the Ph.D. degree at NYU. He worked with Nokia Bell Labs during the summer of 2018. His research interests include wireless communications and networks.

Rajeev Kumar (Member, IEEE) received the B.Tech. and M.Tech. degrees in electrical engineering from the Indian Institute of Technology Madras, Chennai, India, in 2013, and the Ph.D. degree in electrical engineering from the NYU Tandon School of Engineering in 2020. He worked with Nokia Bell Labs during the summers of 2017 and 2018. He is currently a Senior System Engineer with Qualcomm Incorporated. His research interests include achieving ultra-high reliable and low latency communications in to 5G and beyond cellular networks.

Pei Liu (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Xi'an Jiaotong University, China, in 1997 and 2000, respectively, and the Ph.D. degree in electrical and computer engineering from Polytechnic University in 2007. He is currently a Research Assistant Professor with the Department of Electrical and Computer Engineering, NYU Tandon School of Engineering. His research interests include designing and analyzing wireless network protocols with an emphasis on cross-layer optimizations and currently interested in next-generation communication networks and software defined radios.

Shivendra S. Panwar (Fellow, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Massachusetts at Amherst, Amherst, MA, USA, in 1986. He is currently a Professor with the Department of Electrical and Computer Engineering, NYU Tandon School of Engineering. He is also the Director of the New York State Center for Advanced Technology in Telecommunications (CATT), the Faculty Director and a Co-Founder of the New York City Media Laboratory, and a member of NYU Wireless. He has coauthored a textbook *TCP/IP Essentials: A Lab based Approach* (Cambridge University Press). His research interest includes performance analysis and design of networks. His current research is focused on cross-layer research issues in wireless networks, and multimedia transport over networks. He was a winner of the IEEE Communications Society's Leonard Abraham Prize in 2004, the ICC Best Paper Award in 2016, and the Sony Research Award. He was also co-awarded the Best Paper in 2011 Multimedia Communications Award. He has served as the Secretary for the Technical Affairs Council of the IEEE Communications Society.