Optimizing Cell Sizes for Ultra-Reliable Low-Latency Communications in 5G Wireless Networks

Changcheng Huang
Systems and Computer Engineering
Carleton University
Ottawa, Canada
huang@sce.carleton.ca

Nhat Hieu Le
Systems and Computer Engineering
Carleton University
Ottawa, Canada
nhathieule@cmail.carleton.ca

ABSTRACT

The millimeter-wave (mmWave) band with large antenna arrays and dense base station deployments has become the prime candidate for 5G mobile systems and key enabler for ultra-reliable low-latency communications (URLLC). In this paper, we propose an approach to estimating the optimal cell sizes of 5G networks that support URLLC services by combining both physical and data link layers, leveraging concepts from stochastic geometry and queuing theory. Furthermore, the impacts of the densification of base stations on the average blocking probability, which are of practical interest, are investigated with numerical results. The results show that the signal-to-noise-and-interference ratio (SINR) coverage probability and the average blocking probability achieve optimal values at different cell sizes. Moreover, the differences between the two types of optimal values become more significant with higher SINR thresholds. Our results suggest that traditional SINR-based approach for cell sizing will cause over-provisioning of base stations and significantly higher costs. Specifically, we share the insight that the interactions between SINR at physical layer and retransmission at link layer contribute to varying cost saving.

CCS CONCEPTS

• Hardware → Wireless devices; • Networks → Physical links; Wireless access points, base stations and infrastructure.

KEYWORDS

5G Networks, mmWave, SINR, URLLC

ACM Reference format:

Changcheng Huang and Nhat Hieu Le. 2022. Optimizing Cell Sizes for Ultra-Reliable Low-Latency Communications in 5G Wireless Networks. In *MSWiM 2022: The 25th International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, October 24-28, 2022, Montreal, Canada*. ACM, NewYork, NY, USA, 8 pages.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
MSWiM '22, October 24-28, 2022, Montreal, Canada.
© 2022 Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9479-6/22/10...$15.00.
DOI: https://doi.org/10.1145/3551659.3559056

1 Introduction

URLLC is designed to support a plethora of applications that need extremely high reliability communications within a strictly bounded transmission time, such as smart factory automation, autonomous vehicles, and virtual reality [1]. According to the Third Generation Partnership Project (3GPP), the Quality of Service (QoS) requirements of URLLC traffic in terms of reliability and delay are incredibly stringent. The typical value of delay constraint is below one millisecond and the probability of failure delivery is smaller than 10^{-5} [2]. Thus, URLLC traffic poses unprecedented challenging tasks for both existing and upcoming wireless infrastructure that makes the performance analysis of the network systems that support URLLC services badly needed during the designing and deployment stages.

One of the critical enablers for URLLC traffic is the mmWave signals for the access link due to the abundant spectrum resources and high transmission rate. Recent studies suggest that the distinguishing features of mmWave systems promise to deliver high throughputs for supporting exponentially increasing users who require ubiquitous access to high peak data rates of cellular data [3]. Additionally, a smaller wavelength enables large-dimensional antenna arrays packed at both the sending and receiving terminals. The mmWave networks can support directional beamforming with many antennas to provide array gains compensating the path loss and reducing inter/intra-cell interference. It is without a doubt that the mmWave network system is one of the enablers for URLLC service. Consequently, thoroughly understanding the influence of parameters in the mmWave cellular networks for supporting URLLC services is essential for designing and deploying 5G and future advanced wireless networks.

1.1 Related Work

mmWave channels have distinguishing propagation traits as explored by authors in [4]. First, mmWave bands have an increasing path-loss according to Friis transmission formula. Second, mmWave signals are more sensitive to blockages due to severe penetration loss when it travels through particular materials. Hence, indoor users are unlikely to be covered by outdoor base

stations. Third, research in [4] unveiled that the path loss characteristics of line-of-sight (LoS) and non-line-of-sight (NLoS) links vary considerably due to blockages. Therefore, different features between propagation environments need to be considered to provide a robust system analysis of mmWave cellular networks.

A comprehensive analytical method to evaluate the performance of conventional cellular networks has been established, starting with [5] using stochastic geometry. By modeling the base station locations as Poisson point process (PPP) on the 2D plane, the model provided an asymptotic result of the performance in a real-world cellular system. Unfortunately, the method is not applicable to mmWave systems due to their distinguishing propagation characteristics.

Authors in [6] developed an analytical stochastic geometry solution for a systematic study of mmWave cellular networks where the effects of blockage and different propagation characteristics are incorporated. The SINR distribution was evaluated by assuming base stations (BSs) be spatially distributed according to PPP. Several exciting takeaways were derived from this paper. First, the results show that mmWave networks have higher coverage and capacity if BSs are densely deployed. Moreover, the coverage and rate expressions are functions of antenna geometry, network densification, and desired SINR threshold. Furthermore, the BS density should not exceed an optimal value, after which the performance of mmWave networks degrades.

Authors in [7] and [8] extended mathematical approaches to study the impact of statistical channels and antenna models on the performance of mmWave networks. An extensive directional antenna arrays model in mmWave networks was studied in [7], and coverage analysis was presented in cellular and ad-hoc networks. In addition, the multi-slope path loss model was employed in [8].

Although the research mentioned above explored the performance of the mmWave network with various channel and antenna models, they have not studied the performance under URLLC requirements, which needs to consider both physical and link layers. Several interesting aspects of URLLC traffic were explored in [1], such as the impacts of overheads, decoding probability, interface diversity, and the acquisition of channel state information at the transmitter. Likewise, the physical layer challenges to support URLLC services were surveyed in [9], such as packet structure, fading effects, link budget analysis, and modulation and coding scheme. Besides that, the authors of [10] studied the deployment strategy and suitable modulation and coding scheme to satisfy URLLC's stringent requirements. However, these works merely focus on the conventional cellular system.

The 3GPP standards committee proposed the Orthogonal Frequency Division Multiple Access (OFDMA) frame that adopts scalable numerology to accommodate the demanding QoS requirements of URLLC traffic in [11]. In [12], the authors applied queuing theory and conducted simulations to investigate the design of cellular systems to support URLLC communication. Primarily, they explored the fundamental trade-offs between the reliability, system capacity, and the latency requirement for URLLC packets by introducing *M/M/m/k* and *M/D/m/m* queuing models for users at the cell edge. However, the effects of re-transmissions, decoding probability failure, and a finite block-length packet feature were not considered. Adopting the multi-class queuing model, the authors in [13] proposed a simple one-shot transmission scheme to study how resource provisioning in the time-frequency domain affects the system's reliability. The authors then extended the one-shot model to incorporate the Hybrid Automatic Repeat Request (HARQ) scheme. However, the paper did not consider how the deployment strategies of the cellular network such as cell sizes, antenna model and the target SINR threshold affect the overall reliability of the system supporting URLLC service.

1.2 Our Contributions

This paper is designed to address the issues mentioned above and specifically the issues left over by [13]. The key contributions of this paper can be summarized as follows.

1. Existing publications separated physical and link layers dogmatically. In this paper, we will combine physical and data link layers for a comprehensive evaluation on the blocking probability of an mmWave wireless network in support of URLLC communications. Our proposed method will consider various aspects at the physical layer, such as the density of BSs, antenna model, channel model, etc., as well as the essential factors at the data link layer, including decoding failure, delay, and retransmission attempt. We will focus on modeling the relationship between base station density and average blocking probability. To our best knowledge, our approach has not been explored before.
2. Using our analytical models, we will obtain numerical results on optimal cell sizes under both SINR coverage probability and blocking probability for the closest station. We will show the average cell sizes for achieving the two optimal values are not identical and when we increase the SINR threshold, the differences between them grow further. We will pinpoint these varying differences to the interactions between SINR at physical layer and retransmission at link layer. We will show that the optimal cell sizes derived from the traditional SINR based approach will result in smaller cells than necessary, leading to more BSs being deployed and higher deployment and operational costs to cover the same region.

The rest of the paper is organized as follows. The physical layer models will be covered in Section 2. Section 3 will focus on deriving the equations for calculating blocking probability. Numerical results will be introduced in Section 4. Finally, conclusions will be made in Section 5.

2 Physical Layer Models

2.1 Network Model

We assume that outdoor BSs randomly distribute following a homogeneous PPP $\Phi = \{\boldsymbol{X_i}, i \geq 0\}$ of density λ on the plane, where $\boldsymbol{X_i}$ is the location of the i-th base station. We ignore the impact of indoor base stations due to the high penetration loss nature of mmWave bands. Outdoor User Equipment (UE) follow an independent and stationary PPP Φ_u of density λ_u on a 2D plane. Each user is assumed to be associated with the base station having the smallest path loss. Additionally, with stationary properties of the UE point process, a downlink SINR received by a typical user located at the origin O has an identical distribution of an arbitrary user at any location. The serving base station for a typical user is denoted as $\boldsymbol{X_0}$, $\boldsymbol{R_0} = \|\boldsymbol{X_0}\|$ and $\boldsymbol{R_i} = \|\boldsymbol{X_i}\|, i > 0$ are random variables indicating the link length of the serving and i-th base station to the typical user in the order given.

Due to blockages, base stations are thinned into two independent PPPs corresponding to LoS and NLoS base stations. The probability that a station with the link length r is LoS is defined as $b(r) = e^{-\beta r}$, where β is a constant depending on the geometry and the density of blockage process. Consequently, a station with the link of length r is NLoS with a probability $1 - b(r)$. As a result, the LoS and NLoS base stations are governed by two independent non-homogenous PPPs Φ_L and Φ_N with density functions $b(r)\lambda$ and $(1 - b(r))\lambda$, respectively.

2.2 Channel Model

Since the propagation characters of mmWave frequencies for LoS/NLoS links are not identical, different path loss model parameters are applied. Let α_L and α_N be the path loss exponents for LoS and NLoS links, respectively. The path loss of a link with an arbitrary length r is $l(r) = C_L r^{-\alpha_L}$ when the link is LoS and $l(r) = C_N r^{-\alpha_N}$ when the link is NLoS in which C_L, C_N are the path loss intercepts.

We ignore shadowing effects, and each link is assumed to experience independent Nakagami fading with different parameters N_L and N_N for LoS and NLoS. Let us define $\boldsymbol{H_0}$ and $\boldsymbol{H_i}$ be the small-scale fading random variables for the serving and i-th interference links, thus, $|\boldsymbol{H_0}|^2$ and $|\boldsymbol{H_i}|^2$ are distributed as normalized Gamma variables.

2.2 Antenna Model

In this work, we approximate the actual antenna pattern by the sectored model as [6]. We assume that UE and BSs are equipped with large directional antenna arrays supporting analog beamforming. The maximum directionality gain is obtained by adjusting steering orientation at UE as a receiver and its serving station as a transmitter. The steering angles of interference cells are assumed to be uniformly distributed on the plane. Let $\boldsymbol{G_i}, i > 0$ be the total directivity gain from a typical user to the interference base station, $\boldsymbol{X_i}$. The directivity gain $\boldsymbol{G_i}$ is a discrete random variable that has the probability mass function, $\boldsymbol{G_i} = A_k$ with probability B_k ($k \in \{1,2,3,4\}$), where A_k and B_k are defined in Table 1. M_i, S_i, θ_i are parameters denoting main lobe directivity gain, back lobe gain, and a half-power beamwidth of the main lobe at transmitters or receivers; $C_R = \frac{\theta_R}{2\pi}$ and $C_T = \frac{\theta_T}{2\pi}$, where $i \in \{T, R\}$, are constants. For the serving link, the maximum directivity gain is defined as $G_0 = M_R M_T$.

TABLE 1. Probability Mass Function of $\boldsymbol{G_i}$

k	$\boldsymbol{A_k}$	$\boldsymbol{B_k}$
1	$M_R M_T$	$C_R C_T$
2	$M_R S_T$	$C_R(1 - C_T)$
3	$S_R M_T$	$(1 - C_R)C_T$
4	$S_R S_T$	$(1 - C_R)(1 - C_T)$

Based on this model, the downlinks SINR, $\boldsymbol{\gamma_D}$, can be evaluated as follows [6]:

$$\boldsymbol{\gamma_D} = \frac{|\boldsymbol{H_0}|^2 M_T M_t \, l(\boldsymbol{R_0})}{\sigma^2 + \sum_{i>0:\boldsymbol{X_i}\epsilon\Phi} |\boldsymbol{H_i}|^2 \boldsymbol{D_i} l(\boldsymbol{R_i})}, \quad (1)$$

where σ^2 is the normalized thermal noise power by the transmit power, $\boldsymbol{R_i}$ is a random variable that depicts the length of the interference link from UE to the i-th cell.

2.3 SINR Coverage Probability Analysis

The average cell radius of the network is $R_\lambda = \sqrt{1/\pi\lambda}$, which represents the density of the cellular network and decides the inter-site distance. A small average cell size indicates a high BS density in the network. Given the average cell radius R_λ, the SINR coverage probability $P_C(\gamma, R_\lambda)$ of a mmWave base station is defined as the probability that the downlink SINR is larger than a certain threshold γ.

A mobile station connects with a base station through either LoS or NLoS path. The PDF of the distance from a user to its serving LoS station is [6]

$$\hat{f}_L(r; R_\lambda) = e^{-2R_\lambda^{-2} \int_0^{\psi_L(r)} (1-b(t))tdt} f_L(r), \quad (2)$$

where

$$f_L(r) = 2R_\lambda^{-2} r b(r) e^{-2R_\lambda^{-2} \int_0^r b(t)tdt}, \quad (3)$$

$r > 0$ and $\psi_L(r) = \left(\frac{C_N}{C_L}\right)^{1/\alpha_N} r^{\alpha_L/\alpha_N}$. Analogously, the PDF of the distance to the serving NLoS station is

$$\hat{f}_N(r) = e^{-2R_\lambda^{-2} \int_0^{\psi_N(r)} b(t)tdt} f_N(r), \quad (4)$$

where $r > 0$, $\psi_N(r) = \left(\frac{C_L}{C_N}\right)^{1/\alpha_L} r^{\alpha_N/\alpha_L}$ and

$$f_N(r;R_\lambda) = 2R_\lambda^{-2} r(1-b(r)) e^{-2R_\lambda^{-2}\int_0^r (1-b(t))tdt}. \quad (5)$$

The SINR coverage probability can subsequently be computed [6] as

$$P_C(\gamma,R_\lambda) = P_{C,L}(\gamma,R_\lambda) + P_{C,N}(\gamma,R_\lambda), \quad (6)$$

where $P_{C,S}(.)$ for $S \in \{L,N\}$ is the coverage probability given a user connects with a station in Φ_S, and each term can be evaluated as

$$P_{C,L}(\gamma,R_\lambda) \approx \sum_{n=1}^{N_L} (-1)^{n+1} \binom{N_N}{n} \times \int_0^\infty e^{-\frac{n\sigma^2\eta_L r^{\alpha_L}}{C_L M_R M_T}\gamma - q_1(r,n) - q_2(r,n)} \hat{f}_L(r;R_\lambda)dr, \quad (7)$$

and

$$P_{C,N}(\gamma,R_\lambda) \approx \sum_{n=1}^{N_N} (-1)^{n+1} \binom{N_N}{n} \times \int_0^\infty e^{-\frac{n\sigma^2\eta_N r^{\alpha_N}}{C_N M_R M_T}\gamma - v_1(r,n) - v_2(r,n)} \hat{f}_N(r;R_\lambda)dr, \quad (8)$$

where

$$q_1(r,n) = 2R_\lambda^{-2} \sum_{k=1}^4 B_k \times \int_r^\infty h\left(N_L, \frac{n\eta_L \bar{A}_k r^{\alpha_L}}{N_L t^{\alpha_L}}\gamma\right) b(t)tdt, \quad (9)$$

$$q_2(r,n) = 2R_\lambda^{-2} \sum_{k=1}^4 B_k \times \int_{\psi_L(r)}^\infty h\left(N_N, \frac{nC_N\eta_L \bar{A}_k r^{\alpha_L}}{C_L N_N t^{\alpha_N}}\gamma\right)\left(1-b(t)\right)tdt, \quad (10)$$

$$v_1(r,n) = 2R_\lambda^{-2} \sum_{k=1}^4 B_k \times \int_{\psi_N(r)}^\infty h\left(N_L, \frac{nC_L\eta_N \bar{A}_k r^{\alpha_N}}{C_N N_L t^{\alpha_L}}\gamma\right) b(t)tdt, \quad (11)$$

$$v_2(r,n) = 2R_\lambda^{-2} \sum_{k=1}^4 B_k \times \int_r^\infty h\left(N_N, \frac{n\eta_N \bar{A}_k r^{\alpha_N}}{N_N t^{\alpha_L}}\gamma\right)\left(1-b(t)\right)tdt, \quad (12)$$

and $h(x,y) = 1 - 1/(1+y)^x$. For $S \in \{L,N\}, \eta_S = N_S(N_S!)^{\frac{-1}{N_S}}$, N_s are Nakagami small-scale fading parameters. Additionally, $\bar{A}_k = \frac{A_k}{M_R M_T}, k \in \{1,\dots,4\}$ where A_k and B_k are the probabilities and the directivity gain constants defined in Table I.

Importantly, we can interpret the SINR coverage probability as (1) the probability that an arbitrary UE can achieve the desired SINR, (2) the portion of mobile stations which achieve target SINR on average at any time, or (3) the average fraction of the network plane that is in "coverage" area. Therefore, analyzing the SINR coverage probability is a fundamental step to study the blocking probability, which is discussed in the next section.

3 Blocking Probability Analysis

This section will derive the blocking probability of a typical user experiencing at its associating BS.

3.1 Data Link Model

We will consider a scenario where the total bandwidth available for URLLC and enhanced mobile broadband traffic services at each station is W Hz. We assume that upon arrival by superstition/puncturing framework, URLLC packets are scheduled instantly or dropped, and the new packets do not pre-empt an ongoing URLLC transmission. Additionally, each user receives bursts of packets with active periods b_A and idle periods b_I following exponential distributions. During the busy periods, users wake up, listen, and receive bursts of URLLC packets from its associating station that may include re-transmissions.

Moreover, this research considers downlink transmission of URLLC traffic in a Frequency Division Duplex (FDD) based system with dedicated frequency bands for uplink and downlink. Additionally, HARQ schemes are incorporated so that a URLLC packet is allowed up to M_c transmission attempts. Wireless traffic is susceptible to various factors that can corrupt or lose data in transit. Hence, it is reasonable to assume that a base station will refuse to provide services for users whose SINRs are below a predefined threshold.

3.2 User Classes

Cells in a deployment region form a Voronoi diagram covering all users. Under the assumptions in section 2.1, the number of users falling in each cell follows the Poisson distribution with density $\bar{\lambda} = \frac{\lambda_u}{\lambda}$.

Calculating the blocking probability of a mobile station with continuous SINR values is intractable analytically. The closest solution is the generalized Erlang loss model [14]. In order to apply the model, we have to quantize SINR into discrete values. To this end, we divide the whole SINR range into $\mathcal{C}$ segments $[T_1,T_2),\dots,[T_c,T_{c+1}),\dots,[T_\mathcal{C},\infty)$ and all users whose downlink SINRs are between T_c and T_{c+1} are grouped as class c users with a quantized common SINR γ_c. Without loss of generality, we simply set $\gamma_c = T_c$. The probability that a specific user is assigned into class c can be expressed as

$$\hat{P}(c;R_\lambda) = P_C(T_c,R_\lambda) - P_C(T_{c+1},R_\lambda), \quad (13)$$

where $P_C(\gamma,R_\lambda)$ is calculated in (6) and class 0 users are those whose downlink SINR is below T_1 and are rejected by the serving BS due to unstable connection link.

The average number of users that are assigned into class c users is evaluated as

$$\bar{\lambda}_c(R_\lambda) = \bar{\lambda}\,\hat{P}(c; R_\lambda). \quad (14)$$

3.3 Resource Requirements for URLLC Transmission

A URLLC packet of L information bits destined to a class c user requires r_c channel uses from the serving station in the time-frequency domain to send its codewords. As URLLC packet sizes are considerably tiny, we use the results explored by [13] for the finite block-length regime. The channel uses of class c users with the downlink SINR γ_c for a single transmission is approximated by

$$\bar{r_c} \approx \frac{L}{n(\gamma_c)} + \frac{\left(Q^{-1}(\delta)\right)^2 \bar{n}(\gamma_c)}{2\left(n(\gamma_c)\right)^2} + \frac{\left(Q^{-1}(\delta)\right)^2 \bar{n}(\gamma_c)}{2\left(n(\gamma_c)\right)^2} \sqrt{1 + \frac{4Ln(\gamma_c)}{\bar{n}(\gamma_c)\left(Q^{-1}(\delta)\right)^2}}, \quad (15)$$

where $n(x) = log_2(1+x)$, $Q(.)$ is the Q-function, $\bar{n}(x) = (log_2(e))^2(1-(1+x)^{-2})$ and δ is the decoding probability error.

A URLLC packet of class c users is allocated with a bandwidth of f_c for a period t_c for a single shot transmission. The relationship between the allocated resources in the time-frequency domain with the channel uses is bounded by

$$f_c = \frac{\bar{r_c}}{\varepsilon t_c}, \quad (16)$$

where ε is a scalar depending on the OFDMA frame structure and numerology. The bandwidth requirement for all users' URLLC packets is discussed next.

3.4 Bandwidth Constraint

Each class c user will use the bandwidth of h_c for an active interval b_A and will stay inactive for a period b_I. Assume $B_A = \mathbb{E}[b_A]$ and $B_I = \mathbb{E}[b_I]$ are the average active and idle period, respectively. We shall make an important assumption that while users can leave or join a particular cell simultaneously, the average number of users at a certain period is fixed with the rate that equals the density of the PPP of the number of users falling under the base station. Therefore, the number of active users is governed by the Poisson Process with the arrival rate or the average number of active class c users per millisecond will be calculated as $\lambda_c(R_\lambda) = \frac{B_A}{B_A+B_I}\bar{\lambda}_c(R_\lambda)$.

Let $N_c(t)$ be the number of class c users at time t; thus, we can consider $\lambda_c = \mathbb{E}[N_c(t)]$. Furthermore, the average load of class c users in the closest station is defined as

$$\rho_c(R_\lambda) = \lambda_c(R_\lambda) B_A. \quad (17)$$

The condition of class c users being blocked by their closest station will be

$$f_c + \sum_{c=1}^{\mathcal{C}} f_c N_c(t) > W. \quad (18)$$

3.5 Delay Bound under HARQ

This section will study the delay bound of URLLC packets when incorporating HARQ schemes with multiple re-transmissions. A base station can allow up to M_c re-transmission attempts to class c users. However, the total delay must be within D seconds, and the reliability of the transaction must be at least $1-\Delta$ according to the requirement of URLLC communication. After every transmission, the intended receiver sends one-bit feedback to the base station indicating the success/failure status of the packet decoding process. Suppose that the uplink channel is well provisioned so that there are no scheduling and channel access delays. Therefore, the maximum feedback delay denoted by D_c for class c user includes only the processing and propagation delay at the closest cell. The upper delay bound is given by $\sum_{m=1}^{M_c} t_{c,m} + M_c D_c = D$, where $t_{c,m}$ is the allocated interval in the time domain for the m^{th} re-transmission of a URLLC packet from the closest station to an arbitrary class c user. Furthermore, we assume that the base station assigns the identical interval for each attempt, $t_{c,m} = t_c$. Thus, the transmission interval of class c user is bounded by

$$t_c = \frac{D}{M_c} - D_c \quad (19)$$

3.6 Blocking Probability

The blocking probability of class c user is given in the next lemma following the generalized Erlang model [14].

Lemma 1*:* the blocking probability experienced by a class c UE can be formulated as

$$P_B(c; R_\lambda) = \frac{\sum_{\boldsymbol{n}\epsilon \boldsymbol{S_c}} \prod_{\tilde{c}=1}^{\mathcal{C}} \left(\frac{\rho_{\tilde{c}}(R_\lambda)^{n_{\tilde{c}}}}{n_{\tilde{c}}!}\right)}{\sum_{\boldsymbol{n}\epsilon \boldsymbol{S}} \prod_{\tilde{c}=1}^{\mathcal{C}} \left(\frac{\rho_{\tilde{c}}(R_\lambda)^{n_{\tilde{c}}}}{n_{\tilde{c}}!}\right)}, \quad (20)$$

where the vector $\boldsymbol{n} = (n_1, n_2, \ldots, n_{\mathcal{C}})$ denotes the number of active users for each class, $\boldsymbol{S} = \{\boldsymbol{n} \mid \boldsymbol{f n^T} \le W\}$, $\boldsymbol{f} = (f_1, f_2, \ldots, f_{\mathcal{C}})$ is a vector that includes allocated bandwidths for all classes, and $\boldsymbol{S_c} = \{\boldsymbol{n} \mid \boldsymbol{n} \in S,\ \boldsymbol{n} + \boldsymbol{e_c} \notin S\}$, where $\boldsymbol{e_c}$ is a unit vector with only a non-zero element at c^{th} position.

Next, we compute the average probability that a typical user is rejected by its serving station in the following theorem.

Theorem 1*:* the average blocking probability of an arbitrary UE experiences at the nearest station, is

$$P_B(R_\lambda) = \sum_{c=0}^{\mathcal{C}} P_B(c; R_\lambda)\hat{P}(c; R_\lambda). \quad (21)$$

The proof is straightforward using the law of total probability theorem.

4 Numerical Results

We assumed the mmWave network operate at $V = 28$ GHz, and the bandwidth assigned to each user be $W = 50$ MHz. The LoS and NLoS exponents in the path loss formulas were $\alpha_L = 2$ and $\alpha_N = 4$ and the path loss intercepts were $C_L = C_N = (V/4\pi)^2$. The parameters for Nakagami fading were $N_L = 3$ and $N_N = 2$. The inverse of the average LoS range of the network was $\beta = 1/141.4 \quad m^{-1}$. The antenna patterns were fixed at $(M_T, S_T, \theta_T) = (20\ dB, 0\ dB, 30^o)$ for the transmitter and $(M_R, S_R, \theta_R) = (10\ dB, -10\ dB, 45^o)$ for the receiver.

For each URLLC packet, we assumed the information bits $L = 1000$ bits. The QoS delay requirement $D = 1$ msec with 99.9999% reliability or $\Delta = 10^{-6}$. A UE had the maximum delay, $D_c = 0.125$ msec, $\forall c = 1, \ldots, 10$ to send an acknowledgment packet back to the base station. A base station allowed up to five re-transmission attempts for all user classes, which means $M_c = 5, \forall c = 1, \ldots, 10$. The decoding error probability, δ was chosen to fulfill the reliability requirement such that $\delta^{M_c} \leq \Delta$. The arrival rate of active users for each cell was $\bar{\lambda} = 10$ users/msec, and the average active and idle periods were $B_A = B_I = 1$ msec.

The SINR coverage probability as a function of cell radius and SINR threshold is presented in Fig. 1 which shows that the mmWave SINR coverage for a single station is considerably sensitive to the base station density. mmWave networks typically require small cell radii to achieve sufficient SINR coverage. However, unlike the conventional cellular network, further densifying the network does not necessarily improve SINR values due to the tradeoff between signal and interference strength. While a typical user receives higher serving power corresponding to the numerator in Equation (1), the interference term in the denominator of the SINR equation goes up as well with shorter distance.

The non-monotonic trend of the distribution of SINR coverage probability with base station density suggests a phase transition between the domination of noise and interference terms in the denominator of Equation (1). Decreasing the cell sizes is equivalent to reducing the distance of the serving BS to a typical UE. When the density is low, e.g., from $R_\lambda = 100$ meters to $R_\lambda = 80$ meters, the probability of serving BSs being LoS BSs increases, which provides higher serving power received by a typical user. However, the interference term is not affected significantly as the interferers are still far enough and are less likely to be LoS stations to the typical user.

The network eventually transits from noise-limited to interference-limited regime when there are enough interfering BSs after the network becomes sufficiently dense. When the base station density further grows, there will be enough strong LoS interferers causing the SINR to degrade significantly, illustrated by comparing the distribution at $R_\lambda = 70$ meters and $R_\lambda = 30$ meters. Thus, there is an optimal density value highlighted in blue after which the performance of mmWave systems starts dropping. As shown in Fig. 1, the optimal value of the network density depends on the SINR threshold γ. When $\gamma = -3$ dB, the optimal value is around 70 meters, and the average cell size around 75 meters is the optimal size when $\gamma = 0$ dB.

Fig. 2 shows the average blocking probability versus cell radius. While the average blocking probability in Fig.2 has transition patterns that follow the same general trend of the SINR coverage probability as shown in Fig. 1, the optimal cell sizes are quite different. At $T_1 = -3$ dB and 0 dB respectively, the optimal cell radii for the average blocking probabilities are around 72.5 and 80 meters (highlighted in blue), which account for 3.6% and 6.7% increases compared to the SINR results in Fig.1. This also means that we can reduce the numbers of BSs by 7.2% and 13.4% respectively to cover the same region. This significant saving comes from the fact that our blocking probability model has considered both link and physical layer factors to achieve URLLC.

To shed some light on the cause of these gains, we show the blocking probabilities for different classes under two different SINR thresholds in Fig.3. A base station will always reject mobile stations whose downlink SINRs are smaller than the SINR threshold. These mobile stations are class 0 users defined earlier. Clearly, the event a user belongs to class 0 is the same event the user is not covered by SINR. Although all other classes are covered by SINR, their packets can still be blocked due to link layer constraints as shown in Fig.3. This contributes to the differences between average blocking probabilities and SINR coverage probabilities.

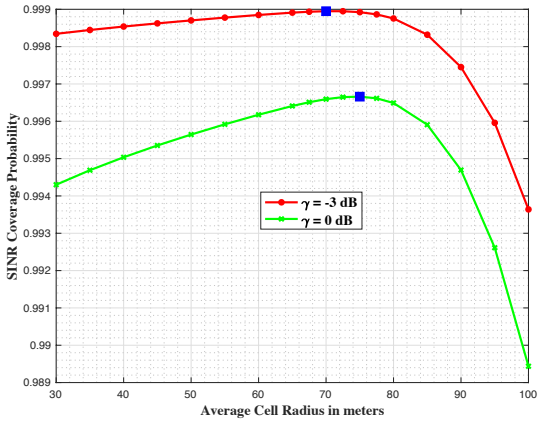

Figure 1: SINR coverage probabilities at various cell radii and different SINR thresholds.

It should be noted that, while class 0 users are not consuming any resources because they are denied access completely, users in other lower classes require more retransmissions than higher classes to satisfy the stringent reliability requirement for URLLC service. As a result, a base station needs to allocate much larger amounts of resources to these lower-class users except class 0.

When SINR thresholds increase beyond the optimal cell sizes predicted by SINR coverage probability, more users close to the

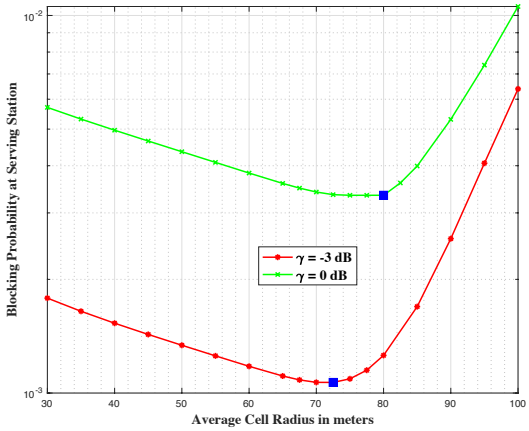

Figure 2: Average blocking probabilities at the closest station with different average cell radii and SINR thresholds.

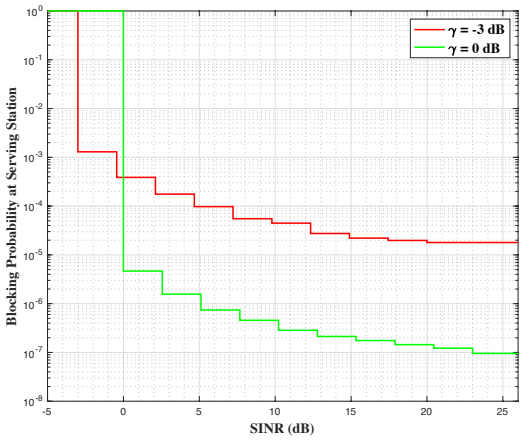

Figure 3: Blocking probability of each class at the closest station with different SINR thresholds at $\boldsymbol{R_\lambda = 100}$m.

edges of the cells will be added to the cells. These users will likely have lower SINR due to weak signal strength. They will join class 0 and tend to drive blocking probabilities higher. However, those users in the cells before the cell sizes are increased will receive less interferences and move to higher classes. They will consume less resources due to less retransmissions. Therefore, they will tend to decrease blocking probabilities and offset the increase introduced by class 0. The overall effect will move the optimal cell sizes to larger values as shown in Fig.2.

When the SINR threshold increases, more lower-class users are rolled over to class 0. While this leads to higher average blocking probability as confirmed in Fig.2, it also significantly reduces the retransmissions dominating in lower classes, making more resources available. Therefore, the blocking probabilities for all other classes go down as illustrated in Fig. 3. This gain in blocking probabilities can somehow partly offset the loss in class 0 users determined by the SINR coverage, shifting the optimal cell sizes to larger values as well.

5 Conclusions

We presented an in-depth study on the blocking probability of URLLC traffic for designing and deploying a 5G wireless network. We focused our attention on the impacts of the parameters at the physical layer over the blocking probability while explicitly considering the effects of delay, retransmissions, and decoding failures at the link layer. In specific, we developed a model that incorporates factors from both link and physical layers. We demonstrated that the optimal blocking probabilities depend on parameters at both physical and link layers through numerical results. Specifically, we have shown simply using SINR as the criterion for deploying 5G wireless networks will lead to over-provisioning of BSs and significantly higher costs. We have also argued that the interactions between SINR thresholds at the physical layer and retransmissions at link layer contribute to the varying gains in optimal cell sizes. This paper also paves the way for further explorations of the impacts of other vital parameters like the antenna or channel model on the performance of the systems that support URLLC service.

REFERENCES

[1] P. Popovski et al, "Ultra-reliable low-latency communication (URLLC): principles and building blocks," *CoRR, vol. abs/1708.07862*, 2017. [Online]. Available: http://arxiv.org/abs/1708.07862.

[2] Chairman's notes 3GPP: 3GPP TSG RAN WG1 Meeting 88bis, Available at http://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_88b/Report/, April 2017.

[3] F. Boccardi et al., "Five Disruptive Technology Directions for 5G," *IEEE Commun. Mag.,* vol. 52, no. 2, pp. 74–80, Feb. 2014.

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

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

[6] T. Bai and R. W. Heath Jr., "Coverage and rate analysis for millimeter-wave cellular networks," *IEEE Trans. Wireless Commun.,* vol. 14, no. 2, pp. 1100–1114, Feb 2015.

[7] X.Yu, J.Zhang, M.Haenggi, and K.B.Letaief, "Coverage analysis for millimeter wave networks:The impact of directional antenna arrays," *IEEE J. on Sel. Areas in Commun.,* vol. 35, no. 7, pp. 1498–1512, Jul. 2017.

[8] X. Zhang and J. G. Andrews, "Downlink cellular network analysis with multi-slope path loss models", *IEEE Trans. Commun.,* vol. 63, no. 5, pp. 1881-1894, May 2015.

[9] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim, "Ultra-reliable and low-latency communications in 5G downlink: Physical layer aspects," *IEEE Wireless Commun.,* vol. 25, no. 3, pp. 124–130, Jun. 2018.

[10] B. Singh, Z. Li, O. Tirkkonen, M. A. Uusitalo, and P. Mogensen, "Ultra-reliable communication in a factory environment for 5G wireless networks: Link level and deployment study," *in IEEE Annual*

Int. Symp. on Personal, Indoor, and Mobile Radio Communications (PIMRC), pp. 1–5, Sept 2016.

[11] 3GPP TR 38.912, "Study on New Radio (NR) access technology," v:15.0.0, Jul. 2018.

[12] C.-P. Li, J. Jiang, W. Chen, T. Ji, and J. Smee, "5G ultra-reliable and low-latency systems design," *in 2017 European Conference on Networks and Communications (EuCNC),* pp. 1–5, June 2017.

[13] A. Anand and G. de Veciana, "Resource Allocation and HARQ Optimization for URLLC Traffic in 5G Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 11, pp. 2411–2421, Nov. 2018.

[14] Hisashi Kobayashi and Brian L Mark. Generalized loss models and queueing-loss networks. *International Transactions in Operational Research*, 9(1): pp. 97–112, 2002.