Downlink Analysis of NOMA-enabled Cellular Networks with 3GPP-inspired User Ranking

Praful D. Mankar, *Member, IEEE,* and Harpreet S. Dhillon, *Senior Member, IEEE*

Abstract—This paper provides a comprehensive downlink analysis of non-orthogonal multiple access (NOMA) enabled cellular networks using tools from stochastic geometry. As a part of this analysis, we develop a novel 3GPP-inspired user ranking technique to construct a user cluster for the non-orthogonal transmission by grouping users from the cell center (CC) and cell edge (CE) regions. This technique allows the pairing of users with distinct link qualities, which is imperative for harnessing NOMA performance gains. The analysis is performed from the perspective of the *typical cell*, which is significantly different from the standard stochastic geometry-based approach of analyzing the performance of the *typical user*. For this setting, we first derive the moments of the meta distributions for the CC and CE users under NOMA and orthogonal multiple access (OMA). Using this, we then derive the distributions of the transmission rates and mean packet delays under non-real time (NRT) and real-time (RT) service models, respectively, for both CC and CE users. Finally, we study two resource allocation (RA) techniques to maximize the cell sum-rate (CSR) under NRT service, and the sum effective capacity (SEC) under RT service. In addition to providing several useful design insights, our results demonstrate that NOMA provides improved rate region and higher CSR as compared to OMA. In addition, we also show that NOMA provides better SEC as compared to OMA for the higher user density.

Index Terms—Non-orthogonal multiple access, user ranking, meta distribution, transmission rate, packet delay, cell sum-rate, effective capacity, Poisson point process.

I. INTRODUCTION

Owing to its superior spectral efficiency, non-orthogonal multiple access (NOMA) technique has emerged as a promising candidate for future wireless cellular networks. Unlike traditional orthogonal multiple access (OMA), NOMA enables the base stations (BSs) to concurrently serve multiple users using the same resource block (RB); see [2] and the references therein. In NOMA, the BS superimposes multiple layers of messages at different power levels and the user decodes its intended message using successive interference cancellation (SIC) technique. In particular, the users with weaker channel qualities are assigned with higher powers so that their intra-cell interference is smaller. For such power allocation, the user first decodes and cancels interference power from the layers assigned to the users with weaker channels successively using SIC and then decodes its intended message. A key design aspect for NOMA is the pairing of users for non-orthogonal transmission. For example, consider there is a user

The authors are with Wireless@VT, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA. Email: {prafuldm, hdhillon}@vt.edu. This paper was presented in part at IEEE Globecom 2019 [1]. The support of the US National Science Foundation (Grant CNS-1814477) is gratefully acknowledged.

with good channel quality, say UE_1, and a user with poor channel quality, say UE_2. In OMA, the associated BS ends up allocating most of the time slots to UE_2 in order to ensure its quality of service (QoS) requirement which leads to lesser transmission opportunities for UE_1 and consequently smaller cell throughput. On the other hand, NOMA allows the BS to concurrently serve both UE_1 and UE_2 using the same spectral resource, which provides uninterrupted channel access to UE_1 while meeting the required QoS constraints of UE_2. Although the UE_1's *per-slot* throughput may suffer (compared to OMA) because of the need to decode both the messages (corresponding to UE_1 and UE_2), the overall transmission rate may still improve because of more transmission opportunities compared to OMA. As a result, NOMA can result in higher cell ergodic capacity as compared to OMA [3]. In this regard, it is beneficial to pair users with distinct link qualities. The readers can refer to [4] for more details on the effects of user pairing on NOMA performance.

A. Prior Art

The set of users scheduled for the non-orthogonal transmission in NOMA is often termed as the *user cluster*. To form a user cluster, one needs to first rank users based on their channel gains (both large-scale path losses and small-scale fading effects) and perceived inter-cell interference, and then place users with distinct link qualities in the same cluster [4]. However, since the prior works on NOMA are mostly focused on the single-cell NOMA analysis, the users are ranked solely based on their distances from the BS [5], [6] or based on their link qualities [3], [7]–[9]. While such user ranking is meaningful in the context of single-cell analysis, it ignores the impact of the inter-cell interference which is crucial for accurate performance analysis of NOMA [10]. Recently, stochastic geometry has emerged as a popular choice for the analysis of large scale cellular networks [11]–[13]. However, incorporating sophisticated user ranking jointly based on the link qualities and inter-cell interference is challenging. This is because of the correlation in the corresponding desired and inter-cell interference powers received by the users within the same cell. Therefore, most of the existing works in this direction ignore these correlations and instead rank the users in the order of their mean desired signal powers (i.e., link distances) so that the i-th closest user becomes the i-th strongest user. Within this general direction, the authors of [14]–[18] analyzed N-ranked NOMA in cellular networks assuming that the BSs follow a PPP. In [18], the uplink success probability is derived assuming that the users follow a

Poisson cluster process (PCP). In [14], the downlink success probability is derived while forming the user cluster within the indisk of the Poisson-Voronoi (PV) cell. However, this may underestimate the NOMA performance gains because users within the indisk of a PV cell will usually experience similar channel conditions and hence lack channel gain imbalance that results in the NOMA gains (see [4]). By ranking the users based on their link distances, the authors of [15], [16] derive the moments of the *meta distribution* of the downlink signal-to-interference ratio (SIR) [19]. However, [16] ignores the joint decoding of the subset of layers associated with SIC. The authors of [15]–[18] use the distribution of the typical link distance (in the network) to derive the order statistics of link distances of clustered users. As implied already, this ignores the correlation in the user locations placed in a PV cell which are a function of the BS point process. A key unintended consequence of this approach is that it does not necessarily confine the user cluster in a PV cell, which is a significant approximation of the underlying setup (see Fig. 1, Middle and Left). The spectral efficiency of the K-tier heterogeneous cellular networks is analyzed in [20] wherein the small cells serve their users using two-user NOMA with distance-based ranking. On similar lines, [21] derives the outage probability for two-user downlink NOMA cellular networks, modeled as a PPP, by ranking the users based on the channel gains normalized by their received inter-cell interference powers. The normalized gains are assumed to be independent and identically distributed (i.i.d.) and follow the distribution that is observed by a typical user in the network. This indeed ignores the correlation in the link distances as well as the inter-cell interference powers associated with the users placed in the same PV cell.

A more reasonable way of accurately ranking the users is to form a user cluster by selecting users from distinct regions of the PV cell. One way of constructing these regions is based on the ratio of mean powers received from the serving and dominant interfering BSs. In particular, a PV cell can be divided into the cell center (CC) region, wherein the ratio is above a threshold τ, and the cell edge (CE) region, wherein the ratio is below τ. A similar approach of classifying users as CC and CE is also used in 3GPP studies to analyze the performance of schemes such as soft frequency reuse (SFR) [22]. Inspired by this, we characterize the CC and CE users based on their path-losses from the serving and dominant interfering BSs to pair them for the two-user NOMA system. While the proposed approach can be directly extended to the N-user NOMA using $N-1$ region partitioning thresholds, we focus on the two-user NOMA for the ease of exposition. The proposed user pairing technique helps to construct the user cluster with distinct link qualities, which is essential for NOMA performance [4]. This is because of two key reasons: 1) the order statistic of received powers at different users in the cell is dominated by their corresponding path-losses [23], and 2) the dominant interfering BS contributes most of the interference power in the PPP setting [24].

Besides user ranking, resource allocation (RA) is an integral part of the NOMA design. On these lines, the prior works have investigated RA for the downlink NOMA system using a variety of performance metrics, such as user fairness [8], [9], [25] and weighted sum-rate maximization [9], [26], [27]. Further, the RA problems for maximizing the sum-rate [9], [14], [28] or the energy efficiency [9], [29], [30] subject to a minimum transmission rate constraint have also been investigated. Besides, [31] has demonstrated that the achievable sum-rate in NOMA is always higher than that of the OMA under minimum transmission rate constraints. These RA formulations are meaningful for the full buffer non-real time (NTR) services, such as file downloading, wherein transmission rates usually determine the QoS. However, they are not suitable for delay-sensitive applications with real time (RT) traffic, such as video streaming and augmented reality, which are becoming even more critical in the context of newly emerging applications of wireless communication. For such applications, it is essential that the RA formulations explicitly include the delay constraints. In this context, the effective capacity (EC), defined in [32] as the maximum achievable arrival rate that satisfies a delay QoS constraint, is analyzed for NOMA in [33]–[35] for the downlink case and in [36], [37] for the uplink case. However, the existing works on the delay analysis of NOMA are relatively sparse. Besides, the works mentioned above on RA with the focus on CSR (i.e., [8], [9], [25], [28]–[31]) and EC (i.e., [33]–[37]) are limited to the single-cell setting, and hence they ignore the impact of inter-cell interference.

B. Contributions

The primary objective of this paper is to enable the accurate downlink NOMA analysis of cellular networks from the perspective of stochastic geometry. As we discussed in Section I-A, the existing stochastic geometry-based NOMA analyses do not capture correlation between signal qualities of the paired users that is induced by the fact that these users are located in the same PV cell. In addition, these analyses also do not explicitly pair users with distinct signal qualities which is crucial for harnessing performance gains using NOMA. The user pairing technique presented in this paper overcomes the above limitations while enabling tractable system-level analysis of downlink NOMA using stochastic geometry. Our approach focuses on the performance of the *typical cell*, which departs significantly from the standard approach of analyzing the performance of the *typical user* in a cellular network that is selected independently of the BS locations [38], [39]. The key contributions of our analysis are briefly summarized below.

1) This paper presents a novel 3GPP-inspired user pairing technique for NOMA to accurately select the CC and CE users with distinct link qualities.

2) We derive approximate yet accurate moments of the meta distributions for the CC and CE users belonging to the typical cell under both NOMA and OMA systems. Besides, we also provide tight beta distribution approximations of the meta distributions.

3) Next, we derive approximate distributions of the mean transmission rates and the upper bounds on the distributions of the mean packet delays of the CC and CE users for both NOMA and OMA systems under the random scheduling scheme.

4) Finally, we present RA formulations for NRT and RT services under both NOMA and OMA systems. For the NRT service, the objective is to maximize CSR such that minimum transmission rates of the CC and CE users are satisfied. For the RT service, the aim is to maximize sum EC (SEC) such that the minimum ECs of CC and CE services are achieved. We present an efficient approach to obtain the near-optimal RAs for these formulations.

5) Our numerical results demonstrate that: a) NOMA is beneficial to both CC and CE users as compared to OMA except when the OMA schedules the CC users for most of the time, b) the proposed near-optimal RA under NRT services provide higher CSR in NOMA as compared to that of the OMA, and c) the proposed near-optimal RA under RT services provide improved SEC in NOMA as compared to that of the OMA at a higher user density.

II. SYSTEM MODEL

A. Network Modeling

We model the locations of BSs using the homogeneous PPP Φ with density λ. This paper considers the strongest mean power-based BS association policy. Hence, the coverage region of the BS at $\mathbf{x} \in \Phi$ becomes the PV cell $V_{\mathbf{x}}$ which is given by $V_{\mathbf{x}} = \{\mathbf{y} \in \mathbb{R}^2 : \|\mathbf{y} - \mathbf{x}\| \leq \|\mathbf{y} - \mathbf{x}'\|, \forall \mathbf{x}' \in \Phi\}$. Let $V_{\mathbf{x}c}$ and $V_{\mathbf{x}e}$ be the CC and CE regions, respectively, of the PV cell $V_{\mathbf{x}}$ corresponding to the BS at $\mathbf{x} \in \Phi$, which are defined as

$$V_{\mathbf{x}c} = \{\mathbf{y} \in V_{\mathbf{x}} : \|\mathbf{y} - \mathbf{x}\| \leq \min_{\mathbf{x}' \in \Phi_{\mathbf{x}}} \tau \|\mathbf{y} - \mathbf{x}'\|\}$$
$$\text{and } V_{\mathbf{x}e} = \{\mathbf{y} \in V_{\mathbf{x}} : \|\mathbf{y} - \mathbf{x}\| > \min_{\mathbf{x}' \in \Phi_{\mathbf{x}}} \tau \|\mathbf{y} - \mathbf{x}'\|\}, \quad (1)$$

where $\Phi_{\mathbf{x}} = \Phi \setminus \{\mathbf{x}\}$ and $\tau \in (0, 1)$ is the boundary threshold. Fig. 1 (Left) depicts the CC and CE regions for $\tau = 0.7$. Now, similar to [40], extending the application of the Type I user point process [41], we define the point processes of the locations of the CC and CE users as

$$\Psi_{cc} = \{U(V_{\mathbf{x}c}; N_{\mathbf{x}c}) : \mathbf{x} \in \Phi\}$$
$$\text{and } \Psi_{ce} = \{U(V_{\mathbf{x}e}; N_{\mathbf{x}e}) : \mathbf{x} \in \Phi\}, \quad (2)$$

respectively, where $U(A; N)$ denotes N points chosen independently and uniformly at random from the set A. Here, $N_{\mathbf{x}c}$ and $N_{\mathbf{x}e}$ are the numbers of CC users in V_{oc} and CE users in V_{oe}, respectively, where μ represents the user density. We assume that $N_{\mathbf{x}c}$ and $N_{\mathbf{x}e}$ follow the zero-truncated Poisson distributions with means $\nu|V_{\mathbf{x}c}|$ and $\nu|V_{\mathbf{x}e}|$, respectively. We refer to $N_{\mathbf{x}c}$ and $N_{\mathbf{x}e}$ as the *CC* and *CE loads* of the BS at $\mathbf{x}$. Table I summarizes the system design variables and performance metrics considered in this paper.

Since by the Slivnyak's theorem, conditioning on a point is the same as adding a point to the PPP, we consider that the nucleus of the *typical cell* of the point process $\Phi \cup \{o\}$ is located at the origin o. Thus, the typical cell becomes $V_o = \{\mathbf{y} \in \mathbb{R}^2 \mid \|\mathbf{y} - \mathbf{x}\| > \|\mathbf{y}\| \forall \mathbf{x} \in \Phi\}$. For this setting, the locations the *typical CC user* and the *typical CE user* of Ψ_{cc} and Ψ_{ce} can be modeled using the uniformly distributed points in V_{oc} and V_{oe}, respectively. Thus, Φ becomes the point process of the interfering BSs that is observed by the typical CC and CE users.

Let $R_o = \|\mathbf{y}\|$ be the *service link distance*, i.e., the distance between the user at $\mathbf{y} \in V_o$ and its serving BS at o. Let $R_d = \|\mathbf{x}_d - \mathbf{y}\|$ be the distance from the user at $\mathbf{y} \in V_o$ to its dominant interfering BS at $\mathbf{x}_d \in \Phi$ where $\mathbf{x}_d = \arg\max_{\mathbf{x} \in \Phi} \|\mathbf{x} - \mathbf{y}\|^{-\alpha}$ and α is the path-loss exponent. Therefore, the definitions, given in (1) and (2), implicitly classify the CC and CE users based on their distances (i.e., path-losses) from their serving and dominant interfering BSs such that the CC user at $\mathbf{y} \in V_{oc}$ has $R_o \leq \tau R_d$ and the CE user at $\mathbf{y} \in V_{oe}$ has $R_o > \tau R_d$. Fig. 1 (Left) illustrates a typical realization of Ψ_{cc} and Ψ_{ce}. From this figure, it is clear that (1) accurately preserves the CC and CE regions wherein the SIR is expected to be higher and lower, respectively. As a comparison, Fig. 1 (Middle) illustrates a realization of a user cluster that results from the distance-based ranking scheme that is employed in [15]–[17]. As is clearly evident from the figure, the user cluster is not confined to the PV cell, which is an unintended consequence of ignoring correlation in the user locations. This can also be verified by comparing the distributions of the ordered distances used in [15]–[17] with those obtained from the simulations. This comparison is given in Fig. 1 (Right) wherein $\tilde{R}_n$ is the link distance of the n-th closest user from the BS.

It should be noted here that the user clustering described in the above is different from the *geographical* clustering of users and BSs that have been recently modeled using PCPs, e.g., see [42], [43]. While the PCPs are useful in capturing the spatial coupling in the locations of the users and BSs, they do not necessarily partition users in different groups based on their QoS as done in the above scheme. Therefore, the proposed technique to form user clusters can apply to more general studies that may require to partition the users based on their QoS experiences, such as soft frequency reuse [22]. Now, we discuss the downlink NOMA transmission for a randomly selected pair of the CC and CE users in the typical cell in the following subsection.

B. Downlink NOMA Transmission based on the Proposed User Pairing

Each BS is assumed to transmit signal superimposed of two layers corresponding to the messages for the CC and CE users forming a user cluster. Henceforth, the layers intended for the CC and CE users are referred to as the $\mathtt{L}_c$ and $\mathtt{L}_e$ layers, respectively. The $\mathtt{L}_c$ and $\mathtt{L}_e$ layers are encoded at power levels of θP and $(1-\theta)P$, respectively, where P is the transmission power per RB and $\theta \in (0, 1)$. Without loss of generality, we assume $P = 1$ (since we ignore thermal noise). Usually, NOMA allocates more power to the weaker user so that it receives smaller intra-cell interference power compared to the desired signal power. Thus, the CC user first decodes the $\mathtt{L}_e$ layer while treating the power assigned to the $\mathtt{L}_c$ layer as interference. After successfully decoding the $\mathtt{L}_e$ layer, the CC user cancels its signal using SIC from the received signal and then decodes the $\mathtt{L}_c$ layer. The SIRs of the typical CC user at $\mathbf{y} \in V_{oc}$ for decoding the $\mathtt{L}_c$ and $\mathtt{L}_e$ layers are

$$\mathtt{SIR}_e = \frac{h_o R_o^{-\alpha}(1-\theta)}{\theta h_o R_o^{-\alpha} + I_\Phi} \text{ and } \mathtt{SIR}_c = \frac{h_o R_o^{-\alpha}\theta}{I_\Phi}, \quad (3)$$

Table I

| System design variables		Performance metrics	
BS point process	Φ	Meta distribution for CC users	$F_c(\cdot,\cdot)$
BS and user densities	λ and ν	Meta distribution for CE users	$F_e(\cdot,\cdot)$
CC and CE users point processes	Ψ_{cc} and Ψ_{ce}	b-th moments of meta distributions	M_b^c and M_b^e
PV cell associated with BS at $\mathbf{x}$	$V_{\mathbf{x}}$	Transmission rates of CC and CE users	R_c and R_e
CC and CE regions of PV cell V_x	$V_{\mathbf{x}c}$ and $V_{\mathbf{x}e}$	Service rates of CC and CE users	μ_c and μ_e
Number of CC and CE users	$N_{\mathbf{x}c}$ and $N_{\mathbf{x}e}$	CDF of CC user's transmission rate	$\mathcal{R}_c(\cdot,\cdot)$
Service link distance	R_o	CDF of CE user's transmission rate	$\mathcal{R}_e(\cdot,\cdot)$
Dominant interfering link distance	R_d	CDF of upper bound on CC user's delay	$\mathcal{D}_c(\cdot,\cdot)$
Boundary threshold for regions	τ	CDF of upper bound on CE user's delay	$\mathcal{D}_e(\cdot,\cdot)$
Transmission layers	$\mathtt{L_c}$ and $\mathtt{L_e}$	Cell sum rate	$\mathrm{CSR_{NOMA}}$
SIR thresholds	β_c and β_e	Effective capacity of CC service	$\mathrm{EC_{NOMA}^c}$
Power allocated to $\mathtt{L_c}$ and $\mathtt{L_e}$	θ and $(1-\theta)$	Effective capacity of CE service	$\mathrm{EC_{NOMA}^e}$
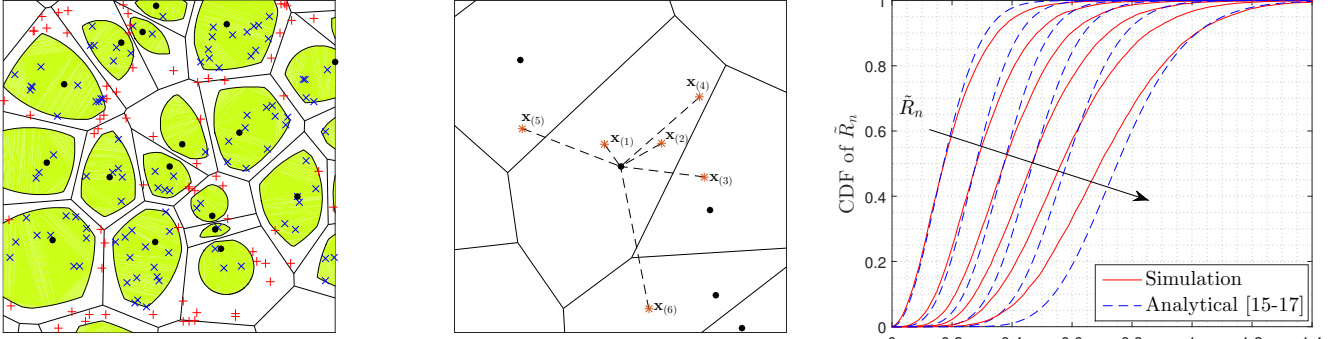
Figure 1. Left: typical realization of Ψ_{cc} and Ψ_{ce} for $\tau = 0.7$, $\lambda = 1$, and $\nu = 20$. Middle: an illustration of the user cluster from [15]–[17] for $N = 6$ (and the fact that the PV cell not necessarily confines the user cluster). Right: the distributions of the ordered link distances modeled in [15]–[17] for $N = 6$ and $\lambda = 1$. The dot, cross, plus, and star markers correspond to BSs, CC users, CE users, and user cluster, respectively. The green and white colors in the left figure show the CC and CE regions.

respectively, where $I_{\Phi} = \sum_{\mathbf{x} \in \Phi} h_{\mathbf{x}} \|\mathbf{x} - \mathbf{y}\|^{-\alpha}$ represents the aggregate inter-cell interference resulting from the set of BSs in Φ and $h_{\mathbf{x}} \sim \exp(1)$ are i.i.d. fading channel gains. Besides, the CE user decodes the $\mathtt{L}_e$ layer while treating the power assigned to the $\mathtt{L}_c$ layer as interference. Thus, the effective SIR of the typical CE user at $\mathbf{y} \in V_{oe}$ is also $\mathtt{SIR}_e$ as given in (3). Note that we assumed a full-buffer system in (3), which is a common assumption in the stochastic geometry literature and is quite reasonable for NRT services for which the objective is to maximize the sum-rate. We will further argue in Subsection II-D that this simple setting also provides useful bounds on the delay-centric metrics.

C. Meta Distribution for the Downlink NOMA System

The success probabilities for the CC and CE users are defined as the probabilities that the typical CC and CE users can decode their intended messages. These success probabilities provide the mean performance of the typical CC and CE users in the network. However, they do not give any information on the disparity in the link performance of the CC and CE users spread across the network. For that purpose, the distribution of the conditional success probability (conditioned on the locations of BS w.r.t. to the CC/CE user location) can be useful. The distribution of the conditional success probability is referred to as the *meta distribution* [19]. The meta distribution for the CC/CE user can be used to answer questions like "what percentage of the CC/CE users can establish their links with the transmission reliability above

a predefined threshold for a given SIR threshold?". Building on the definition of the meta distribution in [19], we define the meta distributions for the CC and CE users under NOMA as below.

Definition 1. *The meta distribution of the typical CC user's success probability is defined as*

$$\bar{F}_c(\beta_c, \beta_e; x) = \mathbb{P}[p_c(\beta_c, \beta_e \mid \mathbf{y}, \Phi) > x], \qquad (4)$$

and the meta distribution of the typical CE user's success probability is defined as

$$\bar{F}_e(\beta_e; x) = \mathbb{P}[p_e(\beta_e \mid \mathbf{y}, \Phi) > x], \qquad (5)$$

where $x \in [0,1]$, β_c and β_e are the SIR thresholds corresponding to the $\mathtt{L}_c$ and $\mathtt{L}_e$ layers, respectively. $p_c(\beta_c, \beta_e \mid \mathbf{y}, \Phi) = \mathbb{P}[\mathtt{SIR}_c \geq \beta_c, \mathtt{SIR}_e \geq \beta_e \mid \mathbf{y}, \Phi]$ and $p_e(\beta_e \mid \mathbf{y}, \Phi) = \mathbb{P}[\mathtt{SIR}_e \geq \beta_e \mid \mathbf{y}, \Phi]$ are the success probabilities of the typical CC and CE users conditioned on their locations at $\mathbf{y}$ and the point process Φ of the interfering BSs, respectively.

D. Traffic Modeling, Scheduling, and Performance Metrics

For the above setting, we will perform a comprehensive load-aware performance analysis of the CC/CE users under the NOMA system. For this, we consider *random scheduling* wherein each BS randomly selects a pair of CC and CE users within its PV cell for the NOMA transmission in a given time slot. For the NRT services, our objective is to maximize the sum-rate. Since the network is assumed to be static, the load-

aware transmission rate of CC/CE user at $\mathbf{y}$ depends on its *scheduling probability* and *successful transmission probability*, both conditioned on Φ. As already implied above, each CC/CE user within a given cell is equally likely to be scheduled in each time slot. Besides, $\mathtt{SIR}$s experienced by the CC/CE user at $\mathbf{y}$ are i.i.d. across the time slots for given Φ. Each BS transmits signal superimposed with $\mathtt{L}_c$ and $\mathtt{L}_e$ layers which are encoded at rates $\mathtt{B}\log_2(1+\beta_c)$ and $\mathtt{B}\log_2(1+\beta_e)$ bits/sec, respectively, where $\mathtt{B}$ is the channel bandwidth. Without loss of generality, here onwards we assume $\mathtt{B}=1$. Therefore, according to Shannon's capacity law, the user requires $\mathtt{SIR}_c$ and $\mathtt{SIR}_e$ above thresholds β_c and β_e, respectively, to successfully decode these layers. Hence, for given Φ, the achievable transmission rates of the CC and CE users located at $\mathbf{y} \in V_o$ under random scheduling respectively become

$$R_c(\mathbf{y}, \Phi) = \frac{p_c(\beta_c, \beta_e \mid \mathbf{y}, \Phi)}{N_{oc}} \log_2(1+\beta_c)$$
$$\text{and } R_e(\mathbf{y}, \Phi) = \frac{p_e(\beta_e \mid \mathbf{y}, \Phi)}{N_{oe}} \log_2(1+\beta_e). \quad (6)$$

The characterization of the typical CC/CE user's success probability under the full-buffer system is also useful in obtaining an upper bound on its delay performance for more general traffic patterns. This can subsequently be used to derive lower bounds on the SEC for the given RT service. The exact delay analysis for the cellular networks is known to be challenging because of the coupled queues at different BSs. Generally, the performance of coupled queues is studied by employing meaningful *modifications* to the system [44]; see [45]–[47] for a small subset of relevant works in the context of cellular networks. For a modified system, the general approach includes the full buffer assumption for the interfering links. As a result, the queue associated with the link of interest operates independently of the statues of queues associated with the interfering links. Note that such a modified system is consistent with our system set-up that is assumed from the very beginning. Since this effectively underestimates the success probability, it provides an upper bound on the packet transmission delay. In the same spirit, this paper presents an upper bound on the distribution of the conditional mean delay of the typical link, which is tighter for the higher load scenarios. One can, of course, determine the mean delay of the typical link by adopting more sophisticated analyses, such as the mean cell approach [48]. However, the contributions of this paper revolve around accurate downlink NOMA analysis using a new user pairing scheme because of which these other approaches are out of the scope of the current paper. We assume that the typical CC/CE user has a dedicated queue of infinite length which is placed at its serving BS. The packet arrival process of the CC/CE service is assumed to follow the Bernoulli distribution with a mean of ϱ_c/ϱ_e packets per slot. The packet sizes of the CC and CE services are considered to be equal to $\mathtt{TB}\log_2(1+\beta_c)$ and $\mathtt{TB}\log(1+\beta_e)$ bits, respectively, where $\mathtt{T}$ is the slot duration. Note that the generalized values of $\mathtt{SIR}$ thresholds β_c and β_e facilitate the choice of selecting CC and CE services with different packet sizes. Thus, the successful packet transmission rates (in packets per slot) for

the typical CC and CE users at $\mathbf{y}$ given Φ respectively become

$$\mu_c(\mathbf{y}, \Phi) = \frac{p_c(\beta_c, \beta_e \mid \mathbf{y}, \Phi)}{N_{oc}}$$
$$\text{and } \mu_e(\mathbf{y}, \Phi) = \frac{p_e(\beta_e \mid \mathbf{y}, \Phi)}{N_{oe}}. \quad (7)$$

Besides, we also analyze the load-aware performances of the CC and CE users under above discussed services for the OMA system. For OMA system, we consider that each BS schedules one of its associated CC (CE) users that is chosen uniformly at random in a given time slot if $r \leq \eta$ ($r > \eta$) where $r \sim U([0,1])$ is generated independently across the time slots. The parameter η allows to control the frequency of scheduling of the CC and CE users in order to meet their QoS requirements. The CC/CE load of the typical cell, and hence the scheduling probability of typical CC/CE user, depends on Φ (see (2)). Thus, from (6)-(7), it is quite evident that the exact analysis requires the joint statistical characterization of the success probability and scheduling probability. However, such joint characterization is challenging as the distribution (even the moments) of the area of the PV cell V_o conditioned on $\mathbf{y} \in V_o$ is difficult to obtain. Hence, similar to [40], [47], we adopt the following reasonable assumption in our analysis.

Assumption 1. *We assume that the CC/CE load (or, the scheduling probability) and the successful transmission probability observed by the typical CC/CE user are independent.*

The numerical results presented in Section VI will demonstrate the accuracy of this assumption for the analysis of the metrics discussed above.

III. META DISTRIBUTION ANALYSIS FOR THE CC AND CE USERS

The main goal of this Section is to present the downlink meta distribution analysis for the NOMA and OMA systems. As stated already in Section I-B, we characterize the performance of the *typical cell* which departs significantly from the standard stochastic geometry approach of analyzing the performance of the typical user. The key intermediate step in the meta distribution analysis is the joint characterization of the service link distance $R_o = \|\mathbf{y}\|$, where $\mathbf{y} \sim U(V_o)$, and the point process Φ of the interfering BSs. Hence, to enable the analysis of the meta distributions of the CC and CE users in the typical cell, we require the joint characterizations of R_o and Φ under the conditions of $R_o \leq R_d\tau$ and $R_o > R_d\tau$. For this, we first determine the marginal and joint probability density functions ($\mathtt{pdf}$s) of R_o and R_d for the CC and CE users. However, given the complexity of the analysis of r.v. R_o [49], it is reasonable to assume that the exact joint characterization of R_o and R_d is equally, if not more, challenging. The marginal distribution of R_o is generally approximated using the contact distribution with adjusted density by a correction factor (c.f.) to maintain tractability [41]. Thus, we approximate the joint $\mathtt{pdf}$ of R_o and R_d using the joint $\mathtt{pdf}$ of the distances to the two nearest points in PPP as

$$f_{R_o, R_d}(r_o, r_d) = (2\pi\rho\lambda)^2 r_o r_d \exp(-\pi\rho\lambda r_d^2), \quad (8)$$

for $r_d \geq r_o \geq 0$, where $\rho = \frac{9}{7}$ is the c.f. (refer [38] for more details).

Lemma 1. *The probabilities that a user uniformly distributed in the typical cell is the CC user and the CE user are equal to τ^2 and $1 - \tau^2$, respectively. The cumulative density function (CDF) of R_o and the CDF of R_d conditioned on R_o for the CC user are given by*

$$F_{R_o}^c(r_o) = 1 - \exp\left(-\pi\rho\lambda r_o^2/\tau^2\right), \qquad (9)$$

for $r_o > 0$, and

$$F_{R_d|R_o}^c\left(r_d \mid r_o\right) = 1 - \exp\left(-\pi\rho\lambda\left(r_d^2 - r_o^2/\tau^2\right)\right), \quad (10)$$

for $r_d > \frac{r_o}{\tau}$, respectively. The joint pdf of R_o and R_d for the CC user is given by

$$f_{R_o,R_d}^c(r_o, r_d) = \frac{(2\pi\rho\lambda)^2}{\tau^2} r_o r_d \exp\left(-\pi\rho\lambda r_d^2\right), \qquad (11)$$

for $r_d > \frac{r_o}{\tau}$ and $r_o > 0$. The CDF of R_o and the CDF of R_d conditioned on R_o for the CE user are respectively given by

$$F_{R_o}^e(r_o) = 1 - \frac{1 - \tau^2 \exp\left(-\pi\rho\lambda r_o^2(\tau^{-2} - 1)\right)}{(1 - \tau^2)\exp\left(\pi\rho\lambda r_o^2\right)}, \qquad (12)$$

for $r_o > 0$, and $F_{R_d|R_o}^e\left(r_d \mid r_o\right)$

$$= \begin{cases} \frac{1 - \exp(-\pi\rho\lambda(r_d^2 - r_o^2))}{1 - \exp(-\pi\rho\lambda r_o^2(\tau^{-2} - 1))}, & \text{for } \frac{r_o}{\tau} > r_d \geq r_o, \\ 1, & \text{for } r_d \geq \frac{r_o}{\tau}. \end{cases} \qquad (13)$$

The joint pdf of R_o and R_d for the CE user is given by

$$f_{R_o,R_d}^e(r_o, r_d) = \frac{(2\pi\rho\lambda)^2}{1 - \tau^2} r_o r_d \exp\left(-\pi\rho\lambda r_d^2\right), \qquad (14)$$

for $\frac{r_o}{\tau} \geq r_d > r_o$ and $r_o > 0$.

Proof. Please refer to Appendix A. $\qquad\square$

A. Meta Distribution for CC and CE users

The CC user needs to decode both the $\mathtt{L}_c$ and $\mathtt{L}_e$ layers for the successful reception its own message. Thus, the successful transmission event for the CC user becomes

$$\begin{aligned} \mathcal{E}_c &= \{\mathtt{SIR}_c > \beta_c\} \cap \{\mathtt{SIR}_e > \beta_e\} \\ &= \{h_o > R_o^\alpha I_\Phi \chi_c\}, \end{aligned} \qquad (15)$$

where $\chi_c = \max\left\{\frac{\beta_c}{\theta}, \frac{\beta_e}{1 - \theta(1 + \beta_e)}\right\}$. On the other hand, the CE user decodes its message while treating the signal intended for the CC user as the interference power. Thus, the successful transmission event for the CE user is given by

$$\mathcal{E}_e = \{\mathtt{SIR}_e > \beta_e\} = \{h_o > R_o^\alpha I_\Phi \chi_e\}, \qquad (16)$$

where $\chi_e = \frac{\beta_e}{1 - \theta(1 + \beta_e)}$. Since it is difficult to derive the meta distribution [19], we first derive its moments in the following theorem and then use them to approximate the meta distribution.

Theorem 1. *The b-th moments of the meta distributions of conditional success probability for the typical CC and CE users under NOMA respectively are*

$$M_b^c(\chi_c) = \frac{\rho^2}{\tau^2} \int_0^{\tau^2} \frac{\left(\rho + v\mathcal{Z}_b\left(\chi_c, v\right)\right)^{-2}}{(1 + \chi_c v^{\frac{1}{\delta}})^b} \mathrm{d}v,$$

$$\text{and } M_b^e(\chi_e) = \frac{\rho^2}{1 - \tau^2} \int_{\tau^2}^1 \frac{\left(\rho + v\mathcal{Z}_b\left(\chi_e, v\right)\right)^{-2}}{(1 + \chi_e v^{\frac{1}{\delta}})^b} \mathrm{d}v, \qquad (17)$$

where $\mathcal{Z}_b(\chi, a) = \chi^\delta \int_{\chi^{-\delta} a^{-1}}^\infty \left[1 - (1 + t^{-\frac{1}{\delta}})^{-b}\right] \mathrm{d}t$ and $\delta = \frac{2}{\alpha}$.

Proof. Please refer to Appendix B. $\qquad\square$

In OMA, each BS serves its associated users using orthogonal RBs which means that there is no intra-cell interference. Thus, OMA provides better success probabilities for the CC and CE users compared to NOMA. However, the orthogonal RB allocation reduces the transmission instances for the CC and CE users, which in turn affects their transmission rates negatively. The successful transmission events for the CC and CE user under OMA respectively given by

$$\tilde{\mathcal{E}}_c = \{h_o > R_o^\alpha \beta_c I_\Phi\} \text{ and } \tilde{\mathcal{E}}_e = \{h_o > R_o^\alpha \beta_e I_\Phi\}. \qquad (18)$$

The following corollary presents the b-th moments of the meta distributions for the OMA case.

Corollary 1. *The b-th moments of the meta distributions of conditional success probability for the typical CC and CE users under OMA respectively are*

$$\tilde{M}_b^c(\beta_c) = \frac{\rho^2}{\tau^2} \int_0^{\tau^2} \frac{\left(\rho + v\mathcal{Z}_b\left(\beta_c, v\right)\right)^{-2}}{(1 + \beta_c v^{\frac{1}{\delta}})^b} \mathrm{d}v,$$

$$\text{and } \tilde{M}_b^e(\beta_e) = \frac{\rho^2}{1 - \tau^2} \int_{\tau^2}^1 \frac{\left(\rho + v\mathcal{Z}_b\left(\beta_e, v\right)\right)^{-2}}{(1 + \chi_e v^{\frac{1}{\delta}})^b} \mathrm{d}v, \qquad (19)$$

where $\mathcal{Z}_b(\beta, a) = \beta^\delta \int_{\beta^{-\delta} a^{-1}}^\infty \left[1 - (1 + t^{-\frac{1}{\delta}})^{-b}\right] \mathrm{d}t$.

Proof. Using the definitions in (18) and following the steps in Appendix B, we obtain (19). $\qquad\square$

Fig. 2 verifies that the means and variances of the meta distributions of the CC and CE users under NOMA (Left) and OMA (Middle) derived in Theorem 1 and Corollary 1, respectively, closely match with the simulation results. The moments for the CE user monotonically decrease with θ as the interference from the $\mathtt{L}_c$ layer increases with θ. However, the performance trend of the moments for the CC user w.r.t. θ is different. This is because θ affects the probabilities of decoding the $\mathtt{L}_c$ and $\mathtt{L}_e$ layers at the CC user differently. While increasing θ makes it difficult to decode $\mathtt{L}_e$ layer, it makes it easier to decode $\mathtt{L}_c$ layer. As a result, the impact of $\mathtt{L}_c$ layer decoding is dominant for $\theta \leq \hat{\theta}$ ($= 0.5$ for $(\beta_c, \beta_e) = (0, -3)$ (in dB)) and the impact of $\mathtt{L}_e$ layer decoding is dominant for $\theta > \hat{\theta}$ where $\hat{\theta}$ will be defined in Section V.
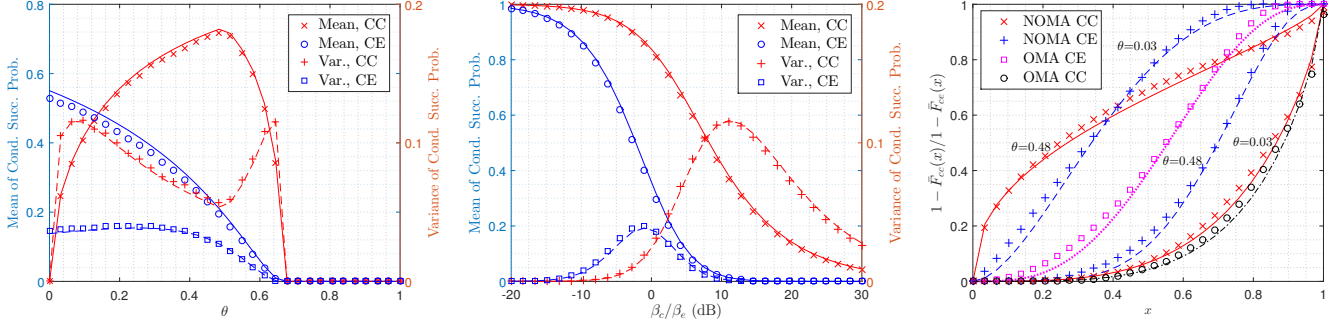
Figure 2. Moments for the CC and CE users under NOMA (Left) and OMA (Middle), and the beta approximations (Right) for $\tau = 0.7$, $\alpha = 4$, $\lambda = 1$, and $(\beta_c, \beta_e) = (0, -3)$ dB. The solid and dashed curves correspond to the analytical results and the markers correspond to the simulation results.

B. Meta Distribution and its Approximation

Using M_b^c and M_b^e derived in Theorem 1 and the Gil-Pelaez's inversion theorem [50], the meta distributions for the typical CC and CE users under NOMA can be obtained. However, the evaluation of meta distributions using the Gil-Pelaez inversion is computationally complex. Thus, similar to [19], we approximate the meta distributions of the CC and CE users with the beta distributions by matching the moments as below

$$\bar{F}_c(\chi_c; x) \approx 1 - I(x; \kappa_{1c}, \kappa_{2c})$$
$$\text{and } \bar{F}_e(\chi_e; x) \approx 1 - I(x; \kappa_{1e}, \kappa_{2e}), \quad (20)$$

respectively, where $I(x; a, b)$ is the regularized incomplete beta function and

$$\kappa_{1s} = \frac{M_1^s \kappa_2^{ss}}{1 - M_1^s} \quad \text{and} \quad \kappa_{2s} = \frac{(M_1^s - M_2^s)(1 - M_1^s)}{M_2^s - (M_1^s)^2}$$

For $s = \{c, e\}$. Similarly, the beta approximations for the meta distributions under OMA can be obtained using the moments given in Corollary 1. We denote the parameters of the beta approximation for the CC (CE) user under OMA by $\tilde{\kappa}_{1c}$ and $\tilde{\kappa}_{1c}$ ($\tilde{\kappa}_{1e}$ and $\tilde{\kappa}_{1e}$). Fig. 2 (Right) shows that the beta distributions closely approximate the meta distributions of the CC and CE users for both NOMA and OMA. The proposed beta approximations can be used for the system-level analysis without having to perform the computationally complex Gil-Pelaez inversion. The figure shows that the percentage of the CE users meeting the link reliability (i.e., conditional success probability) drops with increasing θ whereas the percentage of CC users achieving the link reliability increases with θ.

IV. THROUGHPUT AND DELAY ANALYSIS FOR THE CC AND CE USERS

In this section, we characterize the conditional transmission rate (given in (6)) and the conditional mean delay (given in (7)) of the CC/CE users under NRT and RT services, respectively. For this, we first derive the distributions of the CC and CE loads in the following subsection.

A. Distributions of the CC and CE Loads

The CC and CE loads of the typical cell V_o, i.e., N_{oc} and N_{oc}, depend on the areas of CC and CE regions, i.e., $|V_{oc}|$

and $|V_{oe}|$. It is challenging to derive the area distributions of a random set directly. Thus, we first drive the exact first two moments of these areas in the following lemma and then use them to approximate the area distributions of CC and CE regions.

Lemma 2. *For a given τ, the mean areas of the CC and CE regions are*

$$\mathbb{E}[|V_{oc}|] = \tau^2 \lambda^{-1} \text{ and } \mathbb{E}[|V_{oe}|] = (1 - \tau^2)\lambda^{-1}, \quad (21)$$

respectively, and second moments of the areas of the CC and CE regions are

$$\mathbb{E}[|V_{oc}|^2] = 4\pi \int_0^\pi \int_0^\infty \int_0^\infty \exp(-\lambda U_3) \, r_1 \mathrm{d}r_1 r_2 \mathrm{d}r_2 \mathrm{d}u$$
$$\text{and } \mathbb{E}[|V_{oe}|^2] = 4\pi \int_0^\pi \int_0^\infty \mathcal{F}(r_2, u) r_2 \mathrm{d}r_2 \mathrm{d}u, \quad (22)$$

respectively, where

$$\mathcal{F}(r_2, u) = \int_{\mathtt{D}(r_2, u)} (\exp(-\lambda U_1) \mathbb{1}_{r_1 \le r_2} + \exp(-\lambda U_2) \mathbb{1}_{r_2 < r_1}) \, r_1 \mathrm{d}r_1$$
$$+ \int_{\mathbb{R} \backslash \mathtt{D}(r_2, u)} ([\exp(-\lambda U_o) - \exp(-\lambda(U_1 + U_2 - U_3))] +$$
$$[\exp(-\lambda(U_2 - U_3) - 1][\exp(-\lambda U_1) - \exp(-\lambda U_3)]) r_1 \mathrm{d}r_1,$$

$\mathtt{D}(r_2, u) = \{r_1 \in \mathbb{R} : d \le \tau^{-1}|r_1 - r_2|\}$,
$d = (r_1^2 + r_2^2 - 2r_1 r_2 \cos(u))^{\frac{1}{2}}$, $u_o = u$,
$U_o = U(r_1, r_2, u_o)$, $U_3 = U(r_1 \tau^{-1}, r_2 \tau^{-1}, u_3)$,
$U_1 = U(r_1 \tau^{-1}, r_2, u_1)$ if $r_1 \tau^{-1} < d + r_2$ otherwise $U_1 = \pi r_1^2 \tau^{-2}$,
$U_2 = U(r_1, r_2 \tau^{-1}, u_2)$ if $r_2 \tau^{-1} < d + r_1$ otherwise $U_2 = \pi r_2^2 \tau^{-2}$,
$u_1 = \arccos\left((\tau^{-1} - \tau)\frac{r_1}{2r_2} + \tau \cos(u)\right)$,
$u_2 = \arccos\left((\tau^{-1} - \tau)\frac{r_2}{2r_1} + \tau \cos(u)\right)$,
$u_3 = \arccos\left((1 - \tau^2)\frac{r_1^2 + r_2^2}{2r_1 r_2} + \tau^2 \cos(u)\right)$,
$w(r_1, r_2, u) = \arccos\left(d^{-1}(r_1 - r_2 \cos(u))\right)$, and
$U(r_1, r_2, u) = r_1^2 \left(\pi - w(r_1, r_2, u) + \frac{\sin(2w(r_1, r_2, u))}{2}\right)$
$\quad + r_2^2 \left(\pi - w(r_2, r_1, u) + \frac{\sin(2w(r_2, r_1, u))}{2}\right)$.

Proof. Please refer to Appendix C. □

Now, we approximate the distributions of the CC and CE areas. The area of the PV cell follows the gamma distribution [51] and the sum of two independent gamma distributed random variables also follows the gamma distribution. Therefore, the gamma distribution is the natural choice for these approximations as the correlation between the CC and CE

areas is not too high. Thus, pdfs of the CC and CE areas are, respectively, approximated as

$$f_{|V_{oc}|}(a) = \frac{\gamma_{1c}^{\gamma_{2c}}}{\Gamma(\gamma_{2c})} a^{\gamma_{2c}-1} \exp(-\gamma_{1c}a)$$

$$\text{and } f_{|V_{oe}|}(a) = \frac{\gamma_{1e}^{\gamma_{2e}}}{\Gamma(\gamma_{2e})} a^{\gamma_{2e}-1} \exp(-\gamma_{1e}a), \quad (23)$$

where for $s \in \{c, e\}$ we have

$$\gamma_{2s} = \gamma_{1s}\mathbb{E}[|V_{os}|] \quad \text{and} \quad \gamma_{1s} = \frac{\mathbb{E}[|V_{os}|]}{\mathbb{E}[|V_{os}|^2] - \mathbb{E}[|V_{os}|]^2}.$$

Fig. 3 provides the visual verification of the gamma approximations given in (23). Now, using (23), we obtain the distributions of the CC and CE loads in the following lemma.
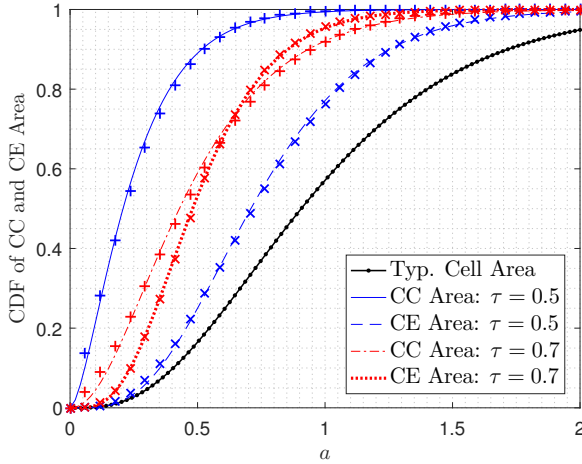

Figure 3. Gamma approximation of the area distributions of the CC and CE regions. The solid and dashed curves correspond to the gamma approximations and the markers correspond to the simulation results.

Lemma 3. *The probability mass function (*pmf*) of the number of CC users, i.e., N_{oc}, is $\mathbb{P}[N_{oc} = n] =$*

$$\frac{\nu^n \gamma_{1c}^{\gamma_{2c}}}{n!\Gamma(\gamma_{2c})} \int_0^\infty a^{n+\gamma_{2c}-1} \frac{\exp(-(\nu + \gamma_{1c})a)}{1 - \exp(-\nu a)} da, \quad (24)$$

where γ_{1c} and γ_{2c} are given in (23). The pmf *of the number of CE, i.e., N_{oe}, is $\mathbb{P}[N_{oe} = n] =$*

$$\frac{\nu^n \gamma_{1e}^{\gamma_{2e}}}{n!\Gamma(\gamma_{2e})} \int_0^\infty a^{n+\gamma_{2e}-1} \frac{\exp(-(\nu + \gamma_{1e})a)}{1 - \exp(-\nu a)} da, \quad (25)$$

where γ_{1e} and γ_{2e} are given in (23).

Proof. For given $|V_{oc}|$, N_{oc} follows zero-truncated Poisson with mean $\nu|V_{oc}|$. Thus, we have

$$\mathbb{P}[N_{oc} = n] = \mathbb{E}_{|V_{oc}|}[\mathbb{P}[N_{oc} = n \mid |V_{oc}|]] \text{ for } n > 0.$$

Now, taking expectation over pdf of $|V_{oc}|$ given in (23) will provide the pmf of N_{oc} as given in (24). Similarly, the pmf of N_{oe} given in (25) follows using the pdf of $|V_{oe}|$ given in (23). □

B. Transmission Rates of the CC and CE Users

In this subsection, we derive the distributions of the conditional transmission rates of the CC and CE users under random scheduling which are, respectively, defined as

$$\mathcal{R}_c(\mathtt{r_c}; \chi_c) = \mathbb{P}[R_c(\mathbf{y}, \Phi) \leq \mathtt{r_c}]$$

$$\text{and } \mathcal{R}_e(\mathtt{r_e}; \chi_e) = \mathbb{P}[R_e(\mathbf{y}, \Phi) \leq \mathtt{r_e}], \quad (26)$$

where $R_c(\mathbf{y}, \Phi)$ and $R_e(\mathbf{y}, \Phi)$ are given in (6). Now, using the meta distributions and the pmfs of the cell loads, the means and the distributions of the conditional transmission rates of the CC and CE users are derived in the following theorem.

Theorem 2. *The mean transmission rates of the typical CC and CE users under NOMA are*

$$\bar{R}_c(\chi_c) = \xi_c \log_2(1 + \beta_c) M_1^c(\chi_c)$$

$$\text{and } \bar{R}_e(\chi_e) = \xi_e \log_2(1 + \beta_e) M_1^e(\chi_e), \quad (27)$$

respectively, where $M_1^c(\chi_c)$ and $M_1^e(\chi_e)$ are given in (17), $s \in \{c, e\}$, $\xi_s = \sum_{n=1}^\infty \frac{1}{n}\mathbb{P}[N_{os} = n]$, and $\mathbb{P}[N_{os} = n]$ is given in Lemma 3. The CDF *of the conditional transmission rates of the typical CC and CE users under NOMA are, respectively,*

$$\mathcal{R}_c(\mathtt{r_c}; \chi_c) =$$
$$\mathbb{E}_{N_{oc}}\left[I\left(\min\left(\mathtt{r_c}N_{oc}\log_2(1 + \beta_c)^{-1}, 1\right); \kappa_{1c}, \kappa_{2c}\right) \right], \quad (28)$$
and $\mathcal{R}_e(\mathtt{r_e}; \chi_e) =$
$$\mathbb{E}_{N_{oe}}\left[I\left(\min\left(\mathtt{r_e}N_{oe}\log_2(1 + \beta_e)^{-1}, 1\right); \kappa_{1e}, \kappa_{2e}\right) \right], \quad (29)$$

where $s \in \{c, e\}$, κ_{1s} and κ_{2s} are given in (20), and $\mathbb{P}[N_{os} = n]$ is given in Lemma 3.

Proof. Please refer to Appendix D. □

For OMA, we consider that each BS serves its associated CC and CE users for η and $1-\eta$ fractions of time, respectively. Note that for $\eta = \frac{|V_{oc}|}{|V_o|}$, the above OMA scheduling scheme will be almost equivalent to the random scheduling wherein the typical BS randomly schedules one of its associated users in a given time slot.

Corollary 2. *The mean transmission rates of the typical CC and CE users under OMA are*

$$\tilde{R}_c(\beta_c) = \eta\xi_c \log_2(1 + \beta_c) \tilde{M}_1^c(\beta_c)$$

$$\text{and } \tilde{R}_e(\beta_e) = (1 - \eta)\xi_e \log_2(1 + \beta_e) \tilde{M}_1^e(\beta_e), \quad (30)$$

respectively, where $\tilde{M}_1^c(\chi_c)$ and $\tilde{M}_1^e(\chi_e)$ are given in (19) and, $s \in \{c, e\}$, $\xi_s = \sum_{n=1}^\infty \frac{1}{n}\mathbb{P}[N_{os} = n]$, and $\mathbb{P}[N_{os} = n]$ is given in Lemma 3. The CDF *of the conditional transmission rates of the typical CC and CE users under OMA respectively are*

$$\tilde{\mathcal{R}}_c(\mathtt{r_c}; \beta_c) =$$
$$\mathbb{E}_{N_{oc}}\left[I\left(\min\left(\frac{\mathtt{r_c}N_{oc}\eta^{-1}}{\log_2(1 + \beta_c)}, 1\right); \tilde{\kappa}_{1c}, \tilde{\kappa}_{2c}\right) \right], \quad (31)$$
and $\tilde{\mathcal{R}}_e(\mathtt{r_e}; \beta_e) =$
$$\mathbb{E}_{N_{oe}}\left[I\left(\min\left(\frac{\mathtt{r_e}N_{oe}(1 - \eta)^{-1}}{\log_2(1 + \beta_e)}, 1\right); \tilde{\kappa}_{1e}, \tilde{\kappa}_{2e}\right) \right], \quad (32)$$

where $s \in \{c, e\}$, $\tilde{\kappa}_{1s}$ and $\tilde{\kappa}_{2s}$ are given in Section III-B, and $\mathbb{P}[N_{os} = n]$ is given in Lemma 3.

Proof. In OMA case, the transmission rates of the CC and CE users, for given $\mathbf{y}$ and Φ, are $\frac{\eta}{N_{oc}}\log_2(1 + \beta_c)\mathbb{E}\left[\mathbb{1}_{\tilde{\mathcal{E}}_c}(\mathtt{SIR}_c) \mid \mathbf{y}, \Phi\right]$ and

$\frac{1-\eta}{N_{oe}} \log_2(1 + \beta_c) \mathbb{E}\left[\mathbb{1}_{\tilde{\mathcal{E}}_e}(\text{SIR}_e) \mid \mathbf{y}, \Phi\right]$, respectively. Hence, further following the steps in Appendix D, we complete the proof. □

C. Delay Analysis of the CC and CE Users

This subsection analyzes the delay performance of the typical CC/CE user for the given RT service under the NOMA setup described in Section II-D. Because of the assumption of saturated queues at the interfering BSs, the meta distributions derived in Section III-A can be directly used to analyze the upper bound of the delay performance of the typical CC and CE user (see Section II-D). That said, the conditional packet transmission rate of the typical CC/CE user is the product of its scheduling probability and success probability as stated in (7). It may be noted that the successful transmission events across the time slots are independent for given $\mathbf{y}$ and Φ. Hence, the service times of packets of the typical CC/CE user at $\mathbf{y}$ given Φ are i.i.d. and follow a geometric distribution with parameter $\mu_c(\mathbf{y}, \Phi)/\mu_e(\mathbf{y}, \Phi)$. Besides, the packet arrives in each time slot as per the Bernoulli process with mean ϱ_c/ϱ_e. Thus, the queue of the typical CC/CE user can be modeled as the Geo/Geo/1 queue. The upper bounds of the conditional mean delays of the typical CC and CE users becomes [52]

$$D_c(\mathbf{y}, \Phi) = \frac{1 - \varrho_c}{\mu_c(\mathbf{y}, \Phi) - \varrho_c} \mathbb{1}_{\mu_c(\mathbf{y}, \Phi) > \varrho_c}$$
$$\text{and } D_e(\mathbf{y}, \Phi) = \frac{1 - \varrho_e}{\mu_e(\mathbf{y}, \Phi) - \varrho_e} \mathbb{1}_{\mu_e(\mathbf{y}, \Phi) > \varrho_e}, \quad (33)$$

respectively. Let $\mathtt{t_c}$ and $\mathtt{t_e}$ are mean delay thresholds of the CC and CE users, respectively.

Theorem 3. *The complementary CDF (CCDF) of the conditional mean delays of the typical CC and CE users under NOMA with random scheduling are upper bounded respectively by*

$$\mathcal{D}_c(\mathtt{t_c}; \chi_c) =$$
$$\mathbb{E}_{N_{oc}}\left[I\left(\min\left(N_{oc}\left(\frac{1 - \varrho_c}{\mathtt{t_c}} + \varrho_c\right), 1\right); \kappa_{1c}, \kappa_{2c}\right)\right], \quad (34)$$
and $\mathcal{D}_e(\mathtt{t_e}; \chi_e) =$
$$\mathbb{E}_{N_{oe}}\left[I\left(\min\left(N_{oe}\left(\frac{1 - \varrho_e}{\mathtt{t_e}} + \varrho_e\right), 1\right); \kappa_{1e}, \kappa_{2e}\right)\right] \quad (35)$$

where $s \in \{c, e\}$, κ_{1s} and κ_{2s} are given in (20), and $\mathbb{P}[N_{oc} = n]$ is given by Lemma 3.

Proof. Using Assumption 1 along with (7) and (33), the CDF of $D_c(\mathbf{y}, \Phi)$ becomes $\mathbb{P}[D_c(\mathbf{y}, \Phi) < \mathtt{t_c}]$

$$= \mathbb{P}\left[\mu_c(\mathbf{y}, \Phi) > \frac{1 - \varrho_c}{\mathtt{t_c}} + \varrho_c, \mu_c(\mathbf{y}, \Phi) > \varrho_c\right],$$
$$= \mathbb{E}_{N_{oc}}\left[\mathbb{P}\left(p_c(\beta_c, \beta_e \mid \mathbf{y}, \Phi) > N_{oc}\left(\frac{1 - \varrho_c}{\mathtt{t_c}} + \varrho_c\right) \mid N_{oc}\right)\right]$$

Further, using the meta distribution for the system model discussed in Section II-D, the CCDF of the upper bounded mean delay of the CC users $D_c(\mathbf{y}, \Phi)$ is given by (34). Similarly, the CCDF of the upper bounded mean delay of CE users $D_e(\mathbf{y}, \Phi)$ is obtained as given in (35). □

For OMA, the mean delay of the typical CC and CE users

at $\mathbf{y}$ conditioned on Φ becomes

$$\tilde{\mu}_c(\mathbf{y}, \Phi) = \frac{\eta \mathbb{E}\left[\mathbb{1}_{\tilde{\mathcal{E}}_c}(\text{SIR}_c) \mid \mathbf{y}, \Phi\right]}{N_{oc}}$$
$$\text{and } \tilde{\mu}_e(\mathbf{y}, \Phi) = \frac{(1 - \eta) \mathbb{E}\left[\mathbb{1}_{\tilde{\mathcal{E}}_e}(\text{SIR}_e) \mid \mathbf{y}, \Phi\right]}{N_{oe}}, \quad (36)$$

The following corollary presents the upper bounded CCDFs of mean delays for the OMA case.

Corollary 3. *The CCDF of the mean delays of the typical CC and CE users under OMA with random scheduling are upper bounded respectively by*

$$\tilde{\mathcal{D}}_c(\mathtt{t_c}; \beta_c) =$$
$$\mathbb{E}_{N_{oc}}\left[I\left(\min\left(\frac{N_{oc}}{\eta}\left(\frac{1 - \varrho_c}{\mathtt{t_c}} + \varrho_c\right), 1\right); \tilde{\kappa}_{1c}, \tilde{\kappa}_{2c}\right)\right], \quad (37)$$

and $\tilde{\mathcal{D}}_e(\mathtt{t_e}; \beta_e) =$
$$\mathbb{E}_{N_{oe}}\left[I\left(\min\left(\frac{N_{oe}}{1 - \eta}\left(\frac{1 - \varrho_e}{\mathtt{t_e}} + \varrho_e\right), 1\right); \tilde{\kappa}_{1e}, \tilde{\kappa}_{2e}\right)\right], \quad (38)$$

where $s \in \{c, e\}$, $\tilde{\kappa}_{1s}$ and $\tilde{\kappa}_{2s}$ are given in Section III-B, and $\mathbb{P}[N_{os} = n]$ is given in Lemma 3.

Proof. Using the successful packet transmission rates $\tilde{\mu}_c(\mathbf{y}, \Phi)$ and $\tilde{\mu}_e(\mathbf{y}, \Phi)$ given in (36) and further following the proof of Theorem 3, we obtained the upper bounds on CCDFs of mean delays of the CC and CE users under OMA as given in (37) and (38), respectively. □

V. RESOURCE ALLOCATION AND PERFORMANCE COMPARISON

In this section, we focus on the RA for maximizing the network performance under both NRT and RT services while meeting the QoS constraints of the CC and CE users. For NRT services, we focus on the maximization of CSR such that the minimum transmission rates of the CC and CE users are ensured. However, for RT services, we consider the maximization of SEC such that CC and CE services with minimum arrival rates can be supported, and their corresponding packet transmission delays are also bounded. First, we develop an efficient method to obtain near-optimal RA for the NRT services in the following subsection.

A. RA under NRT services

Using the success probabilities of the CC and CE users, CSRs for the NOMA and OMA systems can be, respectively, obtained as

$$\text{CSR}_{\text{NOMA}} = \text{B} \log_2(1 + \beta_c) M_1^c(\chi_c)$$
$$+ \text{B} \log_2(1 + \beta_e) M_1^e(\chi_e), \quad (39)$$
$$\text{and } \text{CSR}_{\text{OMA}} = \eta \text{B} \log_2(1 + \beta_c) \tilde{M}_1^c(\beta_c)$$
$$+ (1 - \eta) \text{B} \log_2(1 + \beta_e) \tilde{M}_1^e(\beta_e). \quad (40)$$

The RA formulation for the NRT services is as follows.

- $\mathcal{P}_1$- CSR maximization subject to the minimum mean rates of the CC and CE users.

$$\text{NOMA:} \quad \begin{array}{c} \max \ \text{CSR}_{\text{NOMA}} \\ \text{s.t. } 0 < \theta < 1 \\ \bar{R}_c(\chi_c) \geq \text{R}_c \\ \bar{R}_e(\chi_e) \geq \text{R}_e, \end{array} \quad \text{and OMA:} \quad \begin{array}{c} \max \ \text{CSR}_{\text{OMA}} \\ \text{s.t. } 0 < \eta < 1 \\ \tilde{R}_c(\beta_c) \geq \text{R}_c \\ \tilde{R}_e(\beta_e) \geq \text{R}_e. \end{array}$$

1) Near-optimal RA for $\mathcal{P}_1$-NOMA: It is difficult to obtain an exact optimal RA to $\mathcal{P}_1$-NOMA as it does not fall in the standard convex-optimization framework. Therefore, we present an efficient method to obtain a near-optimal solution based on the insights obtained from the NOMA analysis. It is natural to allocate the remaining power to the CC user after achieving the minimum transmission rate of the CE user in order to maximize the CSR. Therefore, we consider maximizing the mean transmission rate of the CC user under the constraints of the minimum transmission rates of the CC and CE users. One can easily see that $\theta \leq \theta_{\text{NC}} = (1 + \beta_e)^{-1}$ is the necessary condition for $\text{SIR}_e \geq \beta_e$. From (15), we note that the success probability of the CC user increases with the decrease of χ_c. The success probability (thus, the mean transmission rate) of the CC user is an increasing and decreasing functions of θ for $0 < \theta \leq \hat{\theta}$ and $\hat{\theta} < \theta \leq \theta_{\text{NC}}$, respectively, where

$$\hat{\theta} = \operatorname*{argmin}_{0 < \theta \leq \theta_{\text{NC}}} \chi_c = \left\{ \theta : \frac{\beta_c}{\theta} = \frac{\beta_e}{1 - \theta(1 + \beta_e)} \right\}. \quad (41)$$

From the above, we know that there are two solutions (if any exists), denoted by θ_{lc} and θ_{uc}, to $\bar{R}_c(\chi_c) = \text{R}_c$ such that $\theta_{lc} \leq \hat{\theta}$ and $\theta_{uc} \geq \hat{\theta}$. Besides, the transmission rate of the CE user is a non-increasing function of θ. Let θ_e be the solution of $\bar{R}_e(\chi_e) = \text{R}_e$. Hence, from the above discussion, the optimal allocation becomes $\theta^* = \min\{\theta_e, \hat{\theta}\}$ if θ_{lc} and θ_e exist such that $\theta_e \geq \theta_{lc}$.

2) Optimal RA for $\mathcal{P}_1$-OMA: For $\beta_c > \beta_e$ and $\tilde{M}_1^c(\beta_c) > \tilde{M}_1^e(\beta_e)$, one can easily infer that the CSE_{OMA} and mean transmission rate of the CC user are monotonically increasing functions of η. In contrast, the mean transmission rate of the CE user is a monotonically decreasing function of η. Therefore, it is straightforward to choose the optimal solution $\eta^* = \eta_e$ for OMA if $\eta_c \leq \eta_e$ where η_c and η_e are the solutions of $\tilde{R}_c(\chi_c) = \text{R}_c$ and $\tilde{R}_e(\chi_e) = \text{R}_e$, respectively. Note that there is no feasible solution if $\eta_c > \eta_e$.

B. RA under RT services

To evaluate the EC, we define the delay QoS constraint for the CC/CE users as the probability that the delay outage is below O_c/O_e for the given mean delay threshold t_c/t_e. Therefore, using the upper bound of the mean delay outage and the fact that the mean delay outage is a monotonically increasing function of the arrival rate, the lower bounds of ECs for the CC and CE services under NOMA and OMA become

$$\text{EC}_{\text{NOMA}}^s = \{\varrho_s \in \mathbb{R}_+ : \mathcal{D}_s(\text{t}_s, \chi_s) = \text{O}_s\} \quad (42)$$

$$\text{and } \text{EC}_{\text{OMA}}^s = \{\varrho_s \in \mathbb{R}_+ : \tilde{\mathcal{D}}_s(\text{t}_s, \beta_s) = \text{O}_s\}, \quad (43)$$

respectively, where $s = \{c, e\}$. The RAs for RT services are formulated as below.

- $\mathcal{P}_2$- SEC maximization subject to the minimum ECs for the CC and CE users.

$$\text{NOMA:} \quad \begin{array}{c} \max \ \text{EC}_{\text{NOMA}}^c + \text{EC}_{\text{NOMA}}^e \\ \text{s.t. } 0 < \theta < 1 \\ \text{EC}_{\text{NOMA}}^c \geq \bar{\varrho}_c \\ \text{EC}_{\text{NOMA}}^e \geq \bar{\varrho}_e, \end{array} \quad \text{and OMA:} \quad \begin{array}{c} \max \ \text{EC}_{\text{OMA}}^c + \text{EC}_{\text{OMA}}^e \\ \text{s.t. } 0 < \eta < 1 \\ \text{EC}_{\text{OMA}}^c \geq \bar{\varrho}_c \\ \text{EC}_{\text{OMA}}^e \geq \bar{\varrho}_e. \end{array}$$

The EC constraints of RA formulation $\mathcal{P}_2$ can facilitate RT services for the CC and CE users requiring minimum packet arrival rates $\tilde{\varrho}_c$ and $\tilde{\varrho}_e$, respectively. At the same time, this formulation also ensures that the outage probability of the mean delays for the CC (CE) user for the given delay threshold t_c (t_e) is below O_c (O_e). The local mean packet delay (i.e., the mean number of slots required for the successful delivery of packet) is equal to the first inverse moment of the conditional success probability [53]. It is therefore natural that the EC is higher for the lower inverse moment of the conditional success probability. In fact, it will be evident in section VI that the outages of delay and transmission rate follow similar trends w.r.t. the power allocation θ under NOMA or the time allocation η under OMA. Hence, we can infer that the ECs of CC and CE users also behave similar to their transmission rates. Hence, the maximization of SEC is similar to the maximization of CSR.

1) Near-optimal RA for P_2-NOMA: Since $\mathcal{P}_2$-NOMA is an non-convex-optimization problem, we adopt the similar approach developed in subsection V-A to obtain its near-optimal solution. As already implied above, the EC of CC users is an increasing and decreasing functions of θ for $0 < \theta \leq \hat{\theta}$ and $\hat{\theta} < \theta \leq \theta_{\text{NC}}$, respectively. However, the EC for the CE users is a decreasing function of θ. Let $\tilde{\theta}_{lc}$ and $\tilde{\theta}_{uc}$ be the two solutions of $\text{EC}_{\text{NOMA}}^c = \bar{\varrho}_c$ such that $\tilde{\theta}_{lc} \leq \hat{\theta}$ and $\tilde{\theta}_{uc} \geq \hat{\theta}$, and $\tilde{\theta}_e$ is the solution of $\text{EC}_{\text{NOMA}}^e = \bar{\varrho}_e$. Hence, the optimal power allocation becomes $\tilde{\theta}^* = \min(\tilde{\theta}_e, \hat{\theta})$ if $\tilde{\theta}_{lc}$ and $\tilde{\theta}_e$ exist such that $\tilde{\theta}_e \geq \tilde{\theta}_{lc}$.

2) Optimal RA for $\mathcal{P}_2$-OMA: Note that EC_{OMA}^c increases with η as the scheduling probability for the CC users increases with η. However, EC_{OMA}^e follows exactly the opposite trend. Besides, allocating more transmission time to the CC users obviously results in higher SEC because of $\tilde{M}_1^c(\beta_c) > \tilde{M}_1^e(\beta_e)$. Therefore, we can obtain the optimal solution $\tilde{\eta}^* = \tilde{\eta}_e$ for OMA if $\tilde{\eta}_c \leq \tilde{\eta}_e$ where $\tilde{\eta}_c$ and $\tilde{\eta}_e$ are the solutions of $\text{EC}_{\text{OMA}}^c = \bar{\varrho}_c$ and $\text{EC}_{\text{OMA}}^e = \bar{\varrho}_e$, respectively.

C. Performance gain of NOMA

Due to the intra-cell interference, the successful transmission probabilities of both the CC and CE users drop in the NOMA system compared to those in OMA system. However, concurrent transmissions increase their scheduling probabilities in NOMA as compared to those of the OMA system. Since $\tilde{M}_1^c(\beta_c) = M_1^c(\beta_c)$ and $\tilde{M}_1^e(\beta_e) = M_1^c(\beta_e)$, the performance gain in the transmission rates of CC and CE
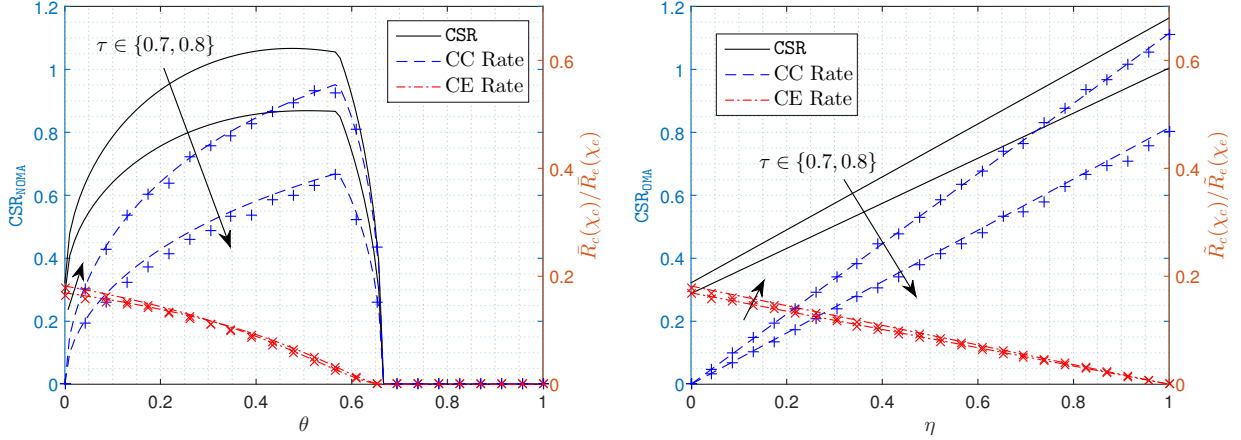
Figure 4. CSR and mean transmission rates of CC and CE users under NOMA (Left) and OMA (Right). The solid and dashed curves correspond to the analytical results, and the markers correspond to the simulation results.

users are, respectively, given by

$$\mathtt{g}_c = \frac{M_1^c(\chi_c)}{\eta M_1^c(\beta_c)} \text{ and } \mathtt{g}_e = \frac{M_1^e(\chi_e)}{(1-\eta)M_1^e(\beta_e)}. \tag{44}$$

Thus, the transmission gain is an increasing function of θ in the interval $(0, \hat{\theta}]$ for the CC users, whereas it is a decreasing function of θ for the CE users. This implies that there is a performance trade-off between the transmission gains for the CC and CE users. For a given η, we have $\mathtt{g}_c > 1$ for a set $\Theta_c(\eta) = \{\theta \in [0, \hat{\theta}] : \frac{M_1^c(\chi_c)}{M_1^c(\beta_c)} > \eta\}$ and $\mathtt{g}_e > 1$ for a set $\Theta_e(\eta) = \{\theta \in [0, \hat{\theta}] : \frac{M_1^e(\chi_e)}{M_1^c(\beta_e)} > 1 - \eta\}$. Therefore, NOMA is beneficial for both CC and CE users compared to OMA only if $\Theta_c(\eta) \cap \Theta_e(\eta) \neq \emptyset$. Otherwise, at least one of these two types of users will underperform in NOMA compared to OMA. It is difficult to analytically show that the intersection of $\Theta_c(\eta)$ and $\Theta(\eta)$ is non-empty for a given η. That said, it will be evident in Section VI that this condition does not hold only for higher values of η.

VI. NUMERICAL RESULTS AND DISCUSSION

In this section, we first verify the accuracy of analytical results by comparing them with the simulation results obtained through Monte Carlo simulations. Next, we discuss the performance trends of the achievable CSR, and the transmission rates and mean delays of the CC and CE users. Further, we also compare the performances of these metrics for the NOMA and OMA systems. For this, we consider $\lambda = 1$, $\nu = 5$, $\alpha = 4$, $\tau = 0.7$, $(\beta_c, \beta_e) = (3, -3)$ in dB, $(\mathtt{r}_c, \mathtt{r}_e) = (0.1, 0.05)$, $(\varrho_c, \varrho_e) = (0.05, 0.05)$, and $(\mathtt{t}_c, \mathtt{t}_e) = (20, 30)$, unless mentioned otherwise.

Fig. 4 depicts that the mean transmission rates of the CC and CE users closely match with the simulation results for both the NOMA and OMA systems. The curves correspond to the analytical results, whereas the markers correspond to the simulation results. The figure gives the visual verifications of the trends of transmission rate functions discussed in Section V. It can be seen that the CSR and the transmission rates are zero for $\theta > \theta_{\mathtt{NC}} \approx 0.66$ which verifies the necessary

condition for NOMA operation. In addition, we can see that the transmission rate of the CE user monotonically decreases with $\theta \leq \theta_{\mathtt{NC}}$, whereas the transmission rate of the CC user is a monotonically increasing and decreasing function of θ for $0 < \theta \leq \hat{\theta}$ and $\hat{\theta} < \theta \leq \theta_{\mathtt{NC}}$, respectively, where $\hat{\theta} = 0.5$ (see (41)).
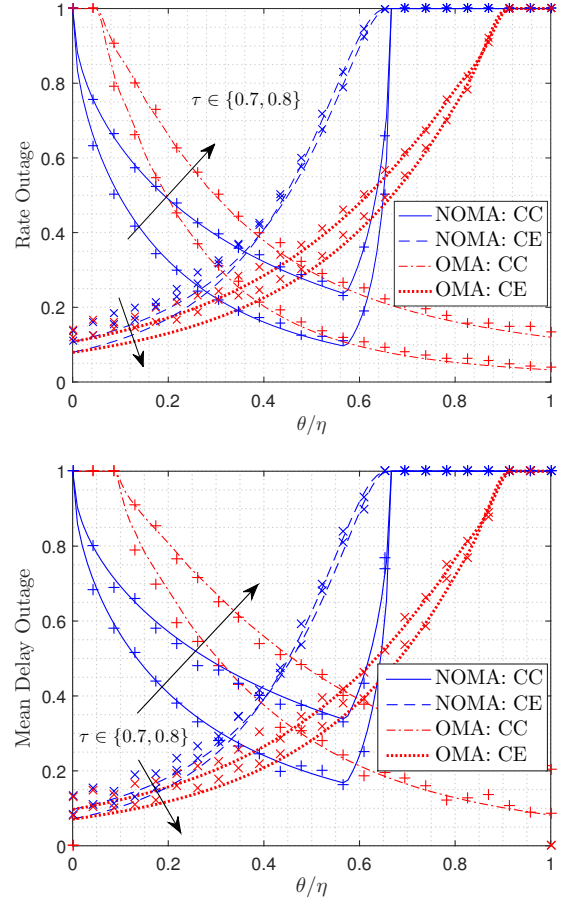

Figure 5. The outage probabilities of transmission rates (Left) and mean delays (Right) of the CC and CE users. The solid and dashed curves correspond to the analytical results and the markers correspond to the simulation results.

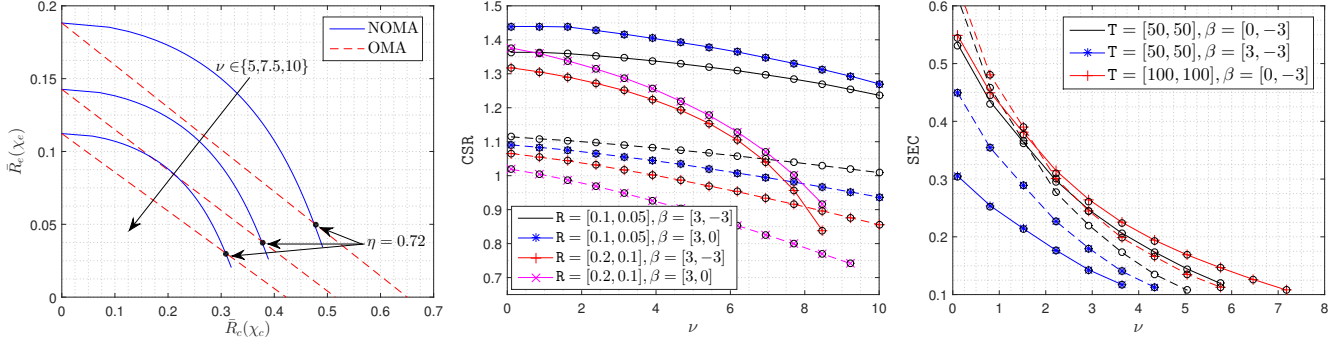Fig. 5 verifies that the transmission rate outage probabilities

Figure 6. Rate region of the CC and CE users in NOMA and OMA for $\beta = [\beta_c, \beta_e] = [3, 0]$ in dB (Left). Maximum CSR under rate constraints where $R = [R_c, R_e]$ (Middle) and maximum SEC under minimum EC constraints (Right) for $(O_c, O_e) = (0.2, 0.2)$ and $(\bar{\varrho}_c, \bar{\varrho}_e) = (0.05, 0.05)$ where $T = [t_c, t_e]$. The solid and dashed lines correspond to the NOMA and OMA, respectively.

and upper bounds of the mean delay outage probabilities of the CC and CE users closely match with the simulation results. The outage probabilities follow the trends opposite to the mean transmission rates. The outage probabilities of the CC user degrade with the increase of τ because of two reasons: 1) increase in the mean number of CC users (i.e., degraded scheduling probability) and 2) decrease in the success probability of CC users. However, a similar direct trend is not visible for the CE user with respect to τ. This is because increasing τ results in the decrease of both the mean number of CE users (i.e., improved scheduling instances) and success probability for the CE users.

Fig. 6 (Left) illustrates the transmission rate region for the CC and CE users under the NOMA and OMA systems. For OMA, the rate region is the linear combination of maximum transmission rates of the CC and CE users where $\eta = 1$ and $\eta = 0$ correspond to their maximum transmission rates, respectively. Figure depicts that the performance gains of both the CC and CE users given in (44) are above unity for $\eta \leq 0.72$ which implies that the intersection of $\Theta_c(\eta)$ and $\Theta_e(\eta)$ are non-empty for $\eta \in [0, 72]$. Further, it can be clearly observed that the NOMA performance gains are independent of ν, whereas the transmission rate regions scale down with the increase of ν. This is because increasing ν lowers the scheduling probabilities of these users, which as a result, affects their transmission rates but not their relative performance gains. Besides, the optimal power allocation with respect to the CC users is $\hat{\theta}$ for the NOMA system. At this point, the CE users receive a non-zero transmission rate as they are allocated with the remaining power of $1 - \hat{\theta}$. Hence, unlike the OMA, the CE users can receive a non-zero transmission rate when the CC users' transmission rate is maximum under the NOMA system.

Fig. 6 (Middle) shows the achievable CSR vs. ν under the minimum transmission rate constraints and Fig 6 (Right) shows the achievable SEC vs. ν under the minimum EC constraints. The circular markers correspond to the proposed solutions and the curves correspond to the exact optimal solutions which are obtained through the brute-force search. We observe that the maximum achievable CSR and SEC drop with the increase of ν for both NOMA and OMA systems. This is because the mean transmission rates of both CC and CE users drop with the increase of ν (i.e., the increase of

the number of users sharing the same RB). This restricts the feasible range of θ and thus that of both CSR and SEC. The middle figure depicts that the NOMA provides better CSR as compared to OMA. The CSR drops at a higher rate when the minimum required transmission rates are higher as it causes the power consumption to increase with the increase in ν aggressively. Fig 6 (Right) depicts that the NOMA provides better SEC for lower SIR threshold at higher values of ν as compared to OMA. This is because their scheduling probabilities predominantly determine the packet service rates of CC and CE users (thus the SEC) for the lower values of SIR thresholds.

VII. Conclusion

This paper has provided a comprehensive analysis of downlink two-user NOMA enabled cellular networks for both RT and NRT services. In particular, a new 3GPP-inspired user ranking technique has been proposed wherein the CC and CE users are paired for the non-orthogonal transmission. To the best of our knowledge, this is the first stochastic geometry-based approach to analyze downlink NOMA using a 3GPP-inspired user ranking scheme that depends upon both the link qualities from the serving and dominant interfering BSs. Unlike the ranking techniques used in the literature, the proposed technique ranks users accurately with distinct link qualities, which is vital to obtain performance gains in NOMA. For the proposed user ranking, we first derive the moments and approximate distributions of the meta distributions for the CC and CE users. Next, we obtain the distributions of their transmission rates and mean delays under random scheduling, which are then used to characterize CSR for NRT service and SEC for RT service, respectively. Using these results, we investigate two RA techniques with objectives to maximize CSR and SEC. The numerical results demonstrated that NOMA, along with the proposed user ranking technique, results in a significantly higher CSR and improved transmission rate region for the CC and CE users as compared to OMA. Besides, our results showed that NOMA provides improved SEC as compared to OMA for the higher user density. The natural extension of this work is to analyze the N-user downlink NOMA in cellular networks. Besides, one could also employ the proposed way of user partitioning to analyze the transmission schemes relying

on the partition of users based on their perceived link qualities, such as SFR.

APPENDIX

A. Proof of Lemma 1

Using (8) and the definitions of Ψ_{cc} and Ψ_{ce} given in (2), it is easy to see that a uniformly distributed user in the typical cell V_o is the typical CC user if $R_o \le \tau R_d$ and the typical CE user if $R_o > \tau R_d$. Therefore, the probability that the typical user is the CC user becomes $\mathbb{P}[R_o \le R_d \tau] =$

$$(2\pi\rho\lambda)^2 \int_0^\infty \int_0^{r_d\tau} r_o r_d \exp(-\pi\lambda\rho r_d^2)\mathrm{d}r_o\mathrm{d}r_d = \tau^2.$$

and thus the probability that the typical user is the CE user becomes $1 - \tau^2$. Using Eq. (8) and $\mathbb{P}[R_o \le R_d\tau] = \tau^2$, we obtain CDF of R_o for the CC user as $F_{R_o}^c(r_o) =$

$$\mathbb{P}[R_o \le r_o | R_o < R_d\tau] = \frac{(2\pi\rho\lambda)^2}{\tau^2} \int_0^{r_o} \int_{\frac{u}{\tau}}^\infty uv \exp(-\pi\rho\lambda v^2)\mathrm{d}v\mathrm{d}u$$
$$= 1 - \exp\left(-\pi\rho\lambda r_o^2/\tau^2\right), \quad (45)$$

for $r_o \ge 0$. Using $R_o \le R_d\tau$, we obtain the CDF of R_d of CC user as

$$F_{R_d|R_o}^c(r_d \mid r_o) = \mathbb{P}[R_d \le r_d \mid R_d > r_o/\tau]$$
$$= 1 - \exp\left(-\pi\rho\lambda\left(r_d^2 - r_o^2/\tau^2\right)\right), \quad (46)$$

for $r_d \ge \frac{r_o}{\tau}$. Therefore, the joint pdf of R_o and R_d for the CC user given in (11) directly follows using the Bayes' theorem along with (45) and (46). Similarly, using (8) and $\mathbb{P}[R_o > R_d\tau] = 1 - \tau^2$, we obtain the CDF of R_o for the CE user as

$$F_{R_o}^e(r_o) = \mathbb{P}[R_o \le r_o \mid R_o > R_d\tau]$$
$$= \frac{(2\pi\rho\lambda)^2}{1 - \tau^2} \int_0^{r_o} \int_u^{\frac{u}{\tau}} uv \exp(-\pi\rho\lambda v^2)\mathrm{d}v\mathrm{d}u$$
$$= 1 - \frac{1 - \tau^2 \exp\left(-\pi\rho\lambda r_o^2(\tau^{-2} - 1)\right)}{(1 - \tau^2)\exp\left(\pi\rho\lambda r_o^2\right)} \quad (47)$$

for $r_o > 0$. Now, the conditional CDF of R_d given R_o for the CE user can be determined as

$$F_{R_d|R_o}^e(r_d \mid r_o) = \mathbb{P}\left[R_d \le r_d \mid R_d < \frac{r_o}{\tau}\right]$$
$$= \frac{\mathbb{P}[R_d \le r_d]}{\mathbb{P}[R_d < \frac{r_o}{\tau}]}$$
$$= \frac{1 - \exp\left(-\pi\rho\lambda(r_d^2 - r_o^2)\right)}{1 - \exp\left(-\pi\rho\lambda r_o^2(\tau^{-2} - 1)\right)}, \quad (48)$$

for $r_o < r_d < \frac{r_o}{\tau}$ and $F_{R_d|R_o}^e(r_d \mid r_o) = 1$ for $r_d \ge \frac{r_o}{\tau}$. Finally, the joint pdf of R_o and R_d for the CE given in Eq. (14) directly follows using the Bayes' theorem and Equations (47) and (48).

B. Proof of Theorem 1

The success probability of the CC user at $\mathbf{y} \in V_o$ conditioned on $\mathbf{y} = R_o$ and Φ is

$$p_c(\chi_c \mid \mathbf{y}, \Phi) = \mathbb{P}(\mathcal{E}_c \mid \mathbf{y}, \Phi) \stackrel{(a)}{=} \prod_{\mathbf{x} \in \Phi} \frac{1}{1 + R_o^\alpha \chi_c \|\mathbf{x} - \mathbf{y}\|^{-\alpha}},$$

where step (a) follows from the independence of the fading gains. Let $\mathbf{x}_d = \arg\min_{\mathbf{x} \in \Phi} \|\mathbf{x} - \mathbf{y}\|$ and $\tilde{\Phi} = \Phi \setminus \{\mathbf{x}_d\}$. Recall $R_d = \|\mathbf{x}_d\|$. The b-th moment of $p_c(\chi_c|\mathbf{y}, \Phi)$ becomes $M_b^c(\chi_c)$

$$= \mathbb{E}_{R_o, \Phi}\left[\prod_{\mathbf{x} \in \Phi} \frac{1}{(1 + R_o^\alpha \chi_c \|\mathbf{x} - \mathbf{y}\|^{-\alpha})^b}\right],$$

$$= \mathbb{E}_{R_o, R_d}\left[\mathbb{E}_{\tilde{\Phi}}\left[\prod_{\mathbf{x} \in \tilde{\Phi}} \frac{(1 + \chi_c(R_o/R_d)^\alpha)^{-b}}{(1 + R_o^\alpha \chi_c \|\mathbf{x} - \mathbf{y}\|^{-\alpha})^b} \mid \mathbf{x}_d\right]\right],$$

$$\stackrel{(a)}{=} \mathbb{E}_{R_o, R_d}\left[\frac{\exp\left(-\lambda \int_{\mathcal{B}_\mathbf{y}(R_d)} \left[1 - (1 + R_o^\alpha \chi_c \|\mathbf{x} - \mathbf{y}\|^{-\alpha})^{-b}\right] \mathrm{d}\mathbf{x}\right)}{(1 + \chi_c(R_o/R_d)^\alpha)^b}\right],$$

$$= \mathbb{E}_{R_o, R_d}\left[\frac{\exp\left(-\pi\lambda R_o^2 \tilde{\mathcal{Z}}_b\left(\chi_c, R_o/R_d\right)\right)}{(1 + \chi_c(R_o/R_d)^\alpha)^b}\right],$$

where $\mathcal{B}_\mathbf{y}^c(r) = \mathbb{R}^2 \setminus \mathcal{B}_\mathbf{y}(r)$, $\tilde{\mathcal{Z}}_b(\chi_c, a) = \chi_c^\delta \int_{\chi_c^{-\delta} a^{-2}}^\infty [1 - (1 + t^{-\frac{1}{\delta}})^{-b}]\mathrm{d}t$, and step (a) follows by approximating $\tilde{\Phi}$ with the homogeneous PPP with density λ outside of the disk $\mathcal{B}_\mathbf{y}(R_d)$ and the probability generating functional of PPP. Now, using the joint pdf of R_o and R_d for the CC user given in (11), we get

$$M_b^c(\chi_c) = \frac{(2\pi\rho\lambda)^2}{\tau^2} \int_0^\infty r_d \exp(-\pi\rho\lambda r_d^2) \times$$
$$\int_0^{\tau r_d} \frac{\exp\left(-\pi\lambda r_o^2 \tilde{\mathcal{Z}}_b\left(\chi_c, r_o/r_d\right)\right)}{(1 + \chi_c(r_o/r_d)^\alpha)^b} r_o \mathrm{d}r_o \mathrm{d}r_d,$$

$$\stackrel{(a)}{=} \frac{2(\pi\rho\lambda)^2}{\tau^2} \int_0^{\tau^2} \frac{1}{(1 + \chi_c v^{\frac{1}{\delta}})^b} \times$$
$$\int_0^\infty r_d^3 \exp\left(-\pi\lambda r_d^2 \left[\rho + \tilde{\mathcal{Z}}_b\left(\chi_c, v^{\frac{1}{2}}\right)\right]\right) \mathrm{d}r_d \mathrm{d}v,$$

where step (a) follows using the substitution $(r_o/r_d)^\alpha = v^{\frac{1}{\delta}}$ and the exchange of the integral orders. Further, solving the inner integral gives $M_b^c(\chi_c)$ as in (17) such that $\mathcal{Z}(\chi_c, v) = \tilde{\mathcal{Z}}(\chi_c, v^{\frac{1}{2}})$. Following similar steps and using the joint pdf of R_o and R_d given in (14), the b-th moment of the conditional success probability for the CE user can be obtained as

$$M_b^e(\chi_e) = \frac{(2\pi\rho\lambda)^2}{1 - \tau^2} \int_0^\infty r_d \exp(-\pi\rho\lambda r_d^2) \times$$
$$\int_{\tau r_d}^{r_d} \frac{\exp\left(-\pi\lambda r_o^2 \tilde{\mathcal{Z}}_b\left(\chi_e, r_o/r_d\right)\right)}{(1 + \chi_e(r_o/r_d)^\alpha)^b} r_o \mathrm{d}r_o \mathrm{d}r_d,$$
$$= \frac{2(\pi\rho\lambda)^2}{1 - \tau^2} \int_{\tau^2}^1 \frac{1}{(1 + \chi_e v^{\frac{1}{\delta}})^b} \times$$
$$\int_0^\infty r_d^3 \exp\left(-\pi\lambda r_d^2 \left[\rho + v\tilde{\mathcal{Z}}_b\left(\chi_e, v^{\frac{1}{2}}\right)\right]\right) \mathrm{d}r_d \mathrm{d}v,$$

Finally, solving the inner integral gives $M_b^e(\chi_e)$ as in (17) where $\mathcal{Z}(\chi_e, v) = \tilde{\mathcal{Z}}(\chi_e, v^{\frac{1}{2}})$.

C. Proof of Lemma 2

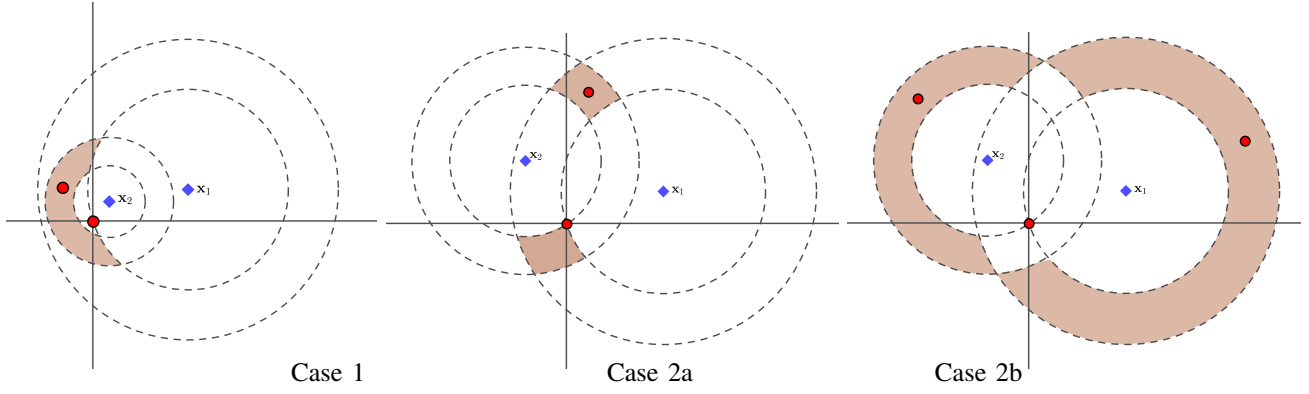The n-th moment of the area of a random set $A \subset \mathbb{R}^2$ can be obtained as [54]

Figure 7. Illustration of the cases when $\{\mathbf{x}_1, \mathbf{x}_2\} \in V_{oe}$ given $\{\mathbf{x}_1, \mathbf{x}_2\} \in V_o$ (i.e. $\Phi(\mathcal{C}_o) = 0$). The blue diamonds represent the locations $\{\mathbf{x}_1, \mathbf{x}_2\}$, whereas the red dots represent the locations of serving and dominant BSs.

$$\mathbb{E}[|A|^n] = \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \mathbb{P}[\mathbf{x}_1, \ldots, \mathbf{x}_n \in A] d\mathbf{x}_1 \ldots d\mathbf{x}_n. \quad (49)$$

Let $\mathcal{B}_\mathbf{x}$ and $\tilde{\mathcal{B}}_\mathbf{x}$ be the disks of radii r and $r\tau^{-1}$ both centered at $\mathbf{x} \equiv (r, \theta)$ and let $\mathcal{A}_\mathbf{x}$ be the annulus formed by the two disks $\tilde{\mathcal{B}}_\mathbf{x}$ and $\mathcal{B}_\mathbf{x}$. By definition, the point $\mathbf{x}$ belongs to V_{oc} only if $\Phi(\mathcal{B}_\mathbf{x}) = 0$ (i.e., $\mathbf{x} \in V_o$) and $\Phi(\mathcal{A}_\mathbf{x}) = 0$. Thus, we have

$$\begin{aligned} \mathbb{P}[\mathbf{x} \in V_{oc}] &= \mathbb{P}[\mathbf{x} \in V_{oc} \mid \mathbf{x} \in V_o]\mathbb{P}[\mathbf{x} \in V_o] \\ &\overset{(a)}{=} \exp\left(-\lambda|\mathcal{A}_\mathbf{x}|\right) \exp(-\lambda|\mathcal{B}_\mathbf{x}|) \\ &= \exp(-\lambda|\tilde{\mathcal{B}}_\mathbf{x}|), \quad (50) \end{aligned}$$

where step (a) follows from the independence property and the *void probability* of the PPP. However, the point $\mathbf{x}$ belongs to V_{oe} only if $\Phi(\mathcal{B}_\mathbf{x}) = 0$ and $\Phi(\mathcal{A}_\mathbf{x}) \neq 0$. Thus, we have

$$\begin{aligned} \mathbb{P}[\mathbf{x} \in V_{oe}] &= \mathbb{P}[\mathbf{x} \in V_{oe} \mid \mathbf{x} \in V_o]\mathbb{P}[\mathbf{x} \in V_o] \\ &\overset{(a)}{=} [1 - \exp\left(-\lambda|\mathcal{A}_\mathbf{x}|\right)]\exp(-\lambda|\mathcal{B}_\mathbf{x}|) \\ &= \exp(-\lambda|\mathcal{B}_\mathbf{x}|) - \exp(-\lambda|\tilde{\mathcal{B}}_\mathbf{x}|), \quad (51) \end{aligned}$$

where step (a) follows from the independence property and the *void probability* of the PPP. Thus, using (49), (50) and (51), we get the mean areas of the CC and CE regions as in (21).

Similarly, the probability of $\{\mathbf{x}_1, \mathbf{x}_2\} \in V_{oc}$ can be directly determined as $\mathbb{P}[\mathbf{x}_1, \mathbf{x}_2 \in V_{oc}] = \exp(-\lambda|\mathcal{C}_3|)$, where $\mathcal{C}_3 = \tilde{\mathcal{B}}_{\mathbf{x}_1} \cup \tilde{\mathcal{B}}_{\mathbf{x}_2}$. Thus, using this and (49), we can easily obtain the second moment of the area of the CC region as in (22). Now, we require the probability of $\{\mathbf{x}_1, \mathbf{x}_2\} \in V_{oe}$ for the evaluation of the second moment of area of the CE region. This requires the careful consideration of the intersection of various sets of two disks. Let $d = \|\mathbf{x}_1 - \mathbf{x}_2\| = (r_1^2 + r_2^2 - 2r_1r_2\cos(\theta_1 - \theta_2))^{\frac{1}{2}}$. Fig. 7 shows two cases wherein Case 1 occurs if $d \leq \tau^{-1}|r_1 - r_2|$, otherwise Case 2 occurs. Now, we derive $\mathbb{P}[\mathbf{x}_1, \mathbf{x}_2 \in V_{oe}]$ for these cases in the following. *Case 1:* In this case, $\{\mathbf{x}_1, \mathbf{x}_2\} \in V_{oe}$ if $\Phi(\mathcal{C}_o) = 0$ (i.e. $\{\mathbf{x}_1, \mathbf{x}_2\} \in V_o$) and if either $\Phi(\mathcal{C}_2 \setminus \mathcal{C}_o) \neq 0$ for $r_2 \leq r_1$ or $\Phi(\mathcal{C}_1 \setminus \mathcal{C}_o) \neq 0$ for $r_1 < r_2$, where $\mathcal{C}_o = \mathcal{B}_{\mathbf{x}_1} \cup \mathcal{B}_{\mathbf{x}_2}$, $\mathcal{C}_1 = \tilde{\mathcal{B}}_{\mathbf{x}_1} \cup \mathcal{B}_{\mathbf{x}_2}$ and $\mathcal{C}_2 = \mathcal{B}_{\mathbf{x}_1} \cup \tilde{\mathcal{B}}_{\mathbf{x}_2}$. Fig. 7 (Left) depicts the second condition of this case for $r_2 \leq r_1$. Thus, we get

$$\mathbb{P}[\mathbf{x}_1, \mathbf{x}_2 \in V_o] = \exp(-\lambda|\mathcal{C}_o|), \quad (52)$$

$$\begin{aligned} \mathbb{P}[\mathbf{x}_1, \mathbf{x}_2 \in V_{oe} \mid \mathbf{x}_1, \mathbf{x}_2 \in V_o] &= \exp\left(-\lambda(|\mathcal{C}_1| - |\mathcal{C}_o|)\right) \mathbb{1}_{r_1 \leq r_2} \\ &+ \exp(-\lambda(|\mathcal{C}_2| - |\mathcal{C}_o|))\mathbb{1}_{r_2 < r_1}. \quad (53) \end{aligned}$$

Therefore, using (52) and (53) and the independence property of the PPP, we obtain $\mathbb{P}[\mathbf{x}_1, \mathbf{x}_2 \in V_{oe}] =$

$$\exp\left(-\lambda|\mathcal{C}_1|\right) \mathbb{1}_{r_1 \leq r_2} + \exp\left(-\lambda|\mathcal{C}_2|\right) \mathbb{1}_{r_2 < r_1}. \quad (54)$$

Case 2: In this case, $\{\mathbf{x}_1, \mathbf{x}_2\} \in V_{oe}$ if $\Phi(\mathcal{C}_o) = 0$ and if one of the following conditions is met: 2a) $\Phi(\mathcal{A}_o) \neq 0$, and 2b) $\Phi(\mathcal{C}_3 \setminus \mathcal{C}_1) \neq 0$, and $\Phi(\mathcal{C}_3 \setminus \mathcal{C}_2) \neq 0$, where $\mathcal{A}_o = \mathcal{A}_{\mathbf{x}_1} \cap \mathcal{A}_{\mathbf{x}_2}$. The above two cases are depicted in Fig. 7 (Middle and Right). Thus, $\mathbb{P}[\mathbf{x}_1, \mathbf{x}_2 \in V_{oe} \mid \mathbf{x}_1, \mathbf{x}_2 \in V_o] =$

$$\begin{aligned} &[1 - \exp(-\lambda|\mathcal{A}_o|)] + \exp(-\lambda|\mathcal{A}_o|) \times \\ &[1 - \exp(-\lambda|\mathcal{C}_3 \setminus \mathcal{C}_1|)][1 - \exp(-\lambda|\mathcal{C}_3 \setminus \mathcal{C}_2|)]. \quad (55) \end{aligned}$$

We have $|\mathcal{A}_o| + |\mathcal{C}_o| = |\mathcal{C}_1| + |\mathcal{C}_2| - |\mathcal{C}_3|$. Therefore, using (52) and (55), we obtain $\mathbb{P}[\mathbf{x}_1, \mathbf{x}_2 \in V_{oe}] =$

$$\begin{aligned} &[\exp(-\lambda(|\mathcal{C}_2| - |\mathcal{C}_3|)) - 1][\exp(-\lambda|\mathcal{C}_1|) - \exp(-\lambda|\mathcal{C}_3|)] \\ &+ [\exp(-\lambda|\mathcal{C}_o|) - \exp(-\lambda(|\mathcal{C}_1| + |\mathcal{C}_2| - |\mathcal{C}_3|))]. \quad (56) \end{aligned}$$

Now, we derive the areas of $\mathcal{C}_o, \mathcal{C}_1, \mathcal{C}_2$, and $\mathcal{C}_3$. Let $U(z_1, z_2, u)$ be the area of the union of two circles of radii z_1 and z_2 with the angular separation of u between their centers w.r.t. their intersection points. Thus, $U(z_1, z_2, u)$ can be easily obtained as given in Lemma 2. Without loss of generality, we set $\theta_2 = 0$ and $0 \leq \theta_1 < \pi$. The two disks in $\mathcal{C}_o, \mathcal{C}_1, \mathcal{C}_2$ and $\mathcal{C}_3$ intersect (if they interest at all) at angles $u_o = \theta_1$,

$$u_1 = \arccos\left((\tau^{-1} - \tau)\frac{r_1}{2r_2} + \tau\cos(\theta_1)\right),$$

$$u_2 = \arccos\left((\tau^{-1} - \tau)\frac{r_2}{2r_1} + \tau\cos(\theta_1)\right)$$

$$\text{and } u_3 = \arccos\left((1 - \tau^2)\frac{r_1^2 + r_2^2}{2r_1r_2} + \tau^2\cos(\theta_1)\right),$$

respectively. The evaluation of $|\mathcal{C}_3|$ is required only for Case 2 wherein $\tilde{\mathcal{B}}_{\mathbf{x}_1}$ and $\tilde{\mathcal{B}}_{\mathbf{x}_2}$ alway intersect. Thus, by definition, we have $|\mathcal{C}_o| = U(r_1, r_2, u_o)$ and $|\mathcal{C}_3| = U\left(r_1\tau^{-1}, r_2\tau^{-1}, u_3\right)$. However, $\tilde{\mathcal{B}}_{\mathbf{x}_1}$ and $\mathcal{B}_{\mathbf{x}_2}$ ($\mathcal{B}_{\mathbf{x}_1}$ and $\tilde{\mathcal{B}}_{\mathbf{x}_2}$) intersect only if $\frac{r_1}{\tau} < d + r_2$ ($\frac{r_2}{\tau} < d + r_1$), otherwise $\mathcal{B}_{\mathbf{x}_2} \subset \tilde{\mathcal{B}}_{\mathbf{x}_1}$ ($\mathcal{B}_{\mathbf{x}_1} \subset \tilde{\mathcal{B}}_{\mathbf{x}_2}$). Thus, we get

$$|\mathcal{C}_1| = \begin{cases} U\left(r_1\tau^{-1}, r_2, u_1\right), & \text{if } r_1\tau^{-1} < d + r_2, \\ \pi r_1^2 \tau^{-2}, & \text{otherwise,} \end{cases}$$

and

$$|\mathcal{C}_2| = \begin{cases} U\left(r_1, r_2\tau^{-1}, u_2\right), & \text{if } r_2\tau^{-1} < d + r_1, \\ \pi r_2^2 \tau^{-2}, & \text{otherwise.} \end{cases}$$

Finally, substituting above expressions in (54) and (56), and then integrating (54) over domain $\{d \le \tau^{-1}|r_1 - r_2|\}$ (i.e., Case 1) and (56) over domain $\{d > \tau^{-1}|r_1 - r_2|\}$ (i.e., Case 2), we get the second moment of the area of the CE region as in (22).

D. Proof of Theorem 2

We first derive the mean transmission rate of the typical CC user at $\mathbf{y} \sim U(V_o)$ as $\bar{R}_c(\chi_c)$

$$= \mathbb{E}_{\mathbf{y},\Phi}\left[\frac{1}{N_{oc}}\log_2(1+\beta_c)\mathbb{E}\left[\mathbb{1}_{\mathcal{E}_c}(\mathtt{SIR}_c, \mathtt{SIR}_e) \mid \mathbf{y}, \Phi\right]\right],$$

$$\overset{(a)}{=} \log_2(1+\beta_c)\mathbb{E}_{|V_{oc}|}\left[\sum_{n=1}^{\infty}\frac{1}{n}\mathbb{P}(N_{oc}=n \mid |V_{oc}|)\right] \times$$
$$\mathbb{E}_{\mathbf{y},\Phi}\left[p_c(\beta_c, \beta_e \mid \mathbf{y}, \Phi)\right],$$

$$= \log_2(1+\beta_c)\sum_{n=1}^{\infty}\frac{1}{n}\mathbb{P}(N_{oc}=n)\mathbb{E}_{\mathbf{y},\Phi}\left[p_c(\beta_c, \beta_e \mid \mathbf{y}, \Phi)\right],$$

where step (a) follows using Assumption 1. From the definition of the meta distribution, we have $\mathbb{E}_{\mathbf{y},\Phi}[p_c(\beta_c, \beta_e \mid \Phi)] = M_1^c(\chi_c, 1)$ for $q_c = q_e = 1$. Hence, using the distribution of CC load given in (24), we obtained $\bar{R}_c(\chi_c)$ as in (27). Similarly, using the distribution of CE load given in (25), we obtained the mean transmission rate $\bar{R}_e(\chi_e)$ of the typical CE user as in (27). Now, we obtain the distribution of the conditional transmission rate of the typical CC user as

$$\mathcal{R}_c(\mathtt{r_c}; \chi_c) = \mathbb{P}\left[p_c(\beta_c, \beta_e \mid \mathbf{y}, \Phi) \le \mathtt{r_c}N_{oc}\log_2(1+\beta_c)^{-1}\right],$$

$$\overset{(a)}{=} \mathbb{E}_{N_{oc}}\left[\mathbb{P}\left[p_c(\beta_c, \beta_e \mid \mathbf{y}, \Phi) \le \mathtt{r_c}N_{oc}\log_2(1+\beta_c)^{-1} \mid N_{oc}\right]\right],$$

$$\overset{(b)}{=} \mathbb{E}_{N_{oc}}\left[I\left(\min\left(\mathtt{r_c}N_{oc}\log_2(1+\beta_c)^{-1}, 1\right); \kappa_{1c}, \kappa_{2c}\right)\right],$$

where step (a) follows using Assumption 1, and step (b) follows using the beta approximation of the meta distribution of the success probability (see (20)). Similarly, we obtain the distribution of the conditional transmission rate of the typical CE user as in (29).

REFERENCES

[1] P. D. Mankar and H. S. Dhillon, "Meta distribution for downlink NOMA in cellular networks with 3GPP-inspired user ranking," *IEEE Globecom*, Dec. 2019.

[2] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

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

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

[5] J. Choi, "Power allocation for max-sum rate and max-min rate proportional fairness in noma," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2055–2058, 2016.

[6] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 938–953, April 2016.

[7] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1462–1465, Aug. 2015.

[8] S. Timotheou and I. Krikidis, "Fairness for non-orthogonal multiple access in 5G systems," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1647–1651, Oct. 2015.

[9] J. Zhu, J. Wang, Y. Huang, S. He, X. You, and L. Yang, "On optimal power allocation for downlink non-orthogonal multiple access systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2744–2757, 2017.

[10] K. S. Ali, H. Elsawy, A. Chaaban, and M.-S. Alouini, "Non-orthogonal multiple access for large-scale 5G networks: Interference aware design," *IEEE Access*, vol. 5, pp. 21204–21216, 2017.

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

[12] J. G. Andrews, A. K. Gupta, and H. S. Dhillon, "A primer on cellular network analysis using stochastic geometry." [Online]. Available: https://arxiv.org/abs/1604.03183

[13] B. Blaszczyszyn, M. Haenggi, P. Keeler, and S. Mukherjee, *Stochastic Geometry Analysis of Cellular Networks*. Cambridge University Press, 2018.

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

[15] K. Ali, H. Elsawy, and M. Alouini, "Meta distribution of downlink non-orthogonal multiple access (NOMA) in Poisson networks," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 572–575, April 2019.

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

[17] ——, "Accuracy of distance-based ranking of users in the analysis of NOMA systems," *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 5069 – 5083, July 2019.

[18] H. Tabassum, E. Hossain, and J. Hossain, "Modeling and analysis of uplink non-orthogonal multiple access in large-scale cellular networks using Poisson cluster processes," *IEEE Trans. Commun.*, vol. 65, no. 8, pp. 3555–3570, Aug. 2017.

[19] M. Haenggi, "The meta distribution of the SIR in Poisson bipolar and cellular networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2577–2589, April 2016.

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

[21] Z. Zhang, H. Sun, and R. Q. Hu, "Downlink and uplink non-orthogonal multiple access in a dense wireless network," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2771–2784, Dec. 2017.

[22] F. Dominique, C. G. Gerlach, N. Gopalakrishnan, A. Rao, J. P. Seymour, R. Soni, A. Stolyar, H. Viswanathan, C. Weaver, and A. Weber, "Self-organizing interference management for LTE," *Bell Labs Tech. J.*, vol. 15, no. 3, pp. 19–42, 2010.

[23] M. Wildemeersch, T. Q. Quek, M. Kountouris, A. Rabbachin, and C. H. Slump, "Successive interference cancellation in heterogeneous networks," *IEEE Trans. Commun.*, vol. 62, no. 12, pp. 4440–4453, Dec. 2014.

[24] V. V. Chetlur and H. S. Dhillon, "Downlink coverage analysis for a finite 3-D wireless network of unmanned aerial vehicles," *IEEE Trans. Commun.*, vol. 65, no. 10, pp. 4543–4558, Oct. 2017.

[25] J. Choi, "Power allocation for max-sum rate and max-min rate proportional fairness in NOMA," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2055–2058, Oct. 2016.

[26] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, "Optimal joint power and subcarrier allocation for full-duplex multicarrier non-orthogonal multiple access systems," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1077–1091, March 2017.

[27] P. Parida and S. S. Das, "Power allocation in OFDM based NOMA systems: A DC programming approach," in *IEEE Globecom Workshops*, Dec. 2014, pp. 1026–1031.

[28] C. Wang, J. Chen, and Y. Chen, "Power allocation for a downlink non-orthogonal multiple access system," *IEEE Wireless Commun. Lett.*, vol. 5, no. 5, pp. 532–535, Oct 2016.

[29] Y. Zhang, H. Wang, T. Zheng, and Q. Yang, "Energy-efficient transmission design in non-orthogonal multiple access," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2852–2857, March 2017.

[30] F. Fang, H. Zhang, J. Cheng, and V. C. Leung, "Energy-efficient resource allocation for downlink non-orthogonal multiple access network," *IEEE Trans. Commun.*, vol. 64, no. 9, pp. 3722–3732, Sept. 2016.

[31] Z. Chen, Z. Ding, X. Dai, and R. Zhang, "An optimization perspective of the superiority of NOMA compared to conventional OMA," *IEEE Trans. Signal Process.*, vol. 65, no. 19, pp. 5191–5202, Oct. 2017.

[32] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 630–643, July 2003.

[33] W. Yu, L. Musavian, and Q. Ni, "Link-layer capacity of NOMA under statistical delay QoS guarantees," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4907–4922, Oct. 2018.

[34] C. Xiao, J. Zeng, W. Ni, X. Su, R. P. Liu, T. Lv, and J. Wang, "Downlink MIMO-NOMA for ultra-reliable low-latency communications," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 780–794, April 2019.

[35] J. Choi, "Effective capacity of NOMA and a suboptimal power control policy with delay QoS," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1849–1858, April 2017.

[36] D. Qiao, M. C. Gursoy, and S. Velipasalar, "Transmission strategies in multiple-access fading channels with statistical QoS constraints," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1578–1593, March 2012.

[37] S. Schiessl, M. Skoglund, and J. Gross, "NOMA in the uplink: Delay analysis with imperfect CSI and finite-length coding." [Online]. Available: https://arxiv.org/abs/1903.09586

[38] P. D. Mankar, P. Parida, H. S. Dhillon, and M. Haenggi, "Downlink analysis for the typical cell in Poisson cellular networks," *IEEE Wireless Commun. Lett.*, Nov. 2019, [Early Access].

[39] P. D. Mankar, H. S. Dhillon, and M. Haenggi, "Meta distribution analysis of the downlink SIR for the typical cell in a Poisson cellular network," *IEEE Globecom*, Dec. 2019.

[40] P. Parida and H. S. Dhillon, "Stochastic geometry-based uplink analysis of massive MIMO systems with fractional pilot reuse," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1651–1668, March 2019.

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

[42] M. Afshang and H. S. Dhillon, "Poisson cluster process based analysis of hetnets with correlated user and base station locations," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2417–2431, April 2018.

[43] P. D. Mankar, G. Das, and S. S. Pathak, "Modeling and coverage analysis of BS-centric clustered users in a random wireless network," *IEEE Wireless Commun. Lett.*, vol. 5, no. 2, pp. 208–211, April 2016.

[44] R. R. Rao and A. Ephremides, "On the stability of interacting queues in a multiple-access system," *IEEE Trans. Inf. Theory*, vol. 34, no. 5, pp. 918–930, Sep. 1988.

[45] T. Bonald, S. Borst, N. Hegde, and A. Proutiére, *Wireless data performance in multi-cell scenarios.* ACM, 2004.

[46] Y. Zhong, M. Haenggi, T. Q. S. Quek, and W. Zhang, "On the stability of static Poisson networks under random access," *IEEE Trans. Commun.*, vol. 64, no. 7, pp. 2985–2998, July 2016.

[47] Y. Zhong, T. Q. S. Quek, and X. Ge, "Heterogeneous cellular networks with spatio-temporal traffic: Delay analysis and scheduling," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1373–1386, June 2017.

[48] B. Bllaszczyszyn, R. Ibrahim, and M. K. Karray, "Spatial disparity of QoS metrics between base stations in wireless cellular networks," *IEEE Trans. Commun.*, vol. 64, no. 10, pp. 4381–4393, Oct. 2016.

[49] P. D. Mankar, P. Parida, H. S. Dhillon, and M. Haenggi, "Distance from the nucleus to a uniformly random point in the 0-cell and the typical cell of the Poisson-Voronoi tessellation." [Online]. Available: https://arxiv.org/abs/1907.03635

[50] J. Gil-Pelaez, "Note on the inversion theorem," *Biometrika*, vol. 38, no. 3-4, pp. 481–482, Dec. 1951.

[51] M. Tanemura, "Statistical distributions of Poisson Voronoi cells in two and three dimensions," *Forma*, vol. 18, no. 4, p. 221–247, 2003.

[52] I. Atencia and P. Moreno, "A discrete-time Geo/G/1 retrial queue with general retrial times," *Queueing systems*, vol. 48, no. 1-2, pp. 5–21, Sept. 2004.

[53] M. Haenggi, "The local delay in Poisson networks," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1788–1802, March 2013.

[54] H. E. Robbins, "On the measure of a random set," *Ann. Math. Statist.*, vol. 15, no. 1, pp. 70–74, March 1944.

Praful D. Mankar (S'14–M'19) Praful received his B.E. degree in Electrical and Communication Engineering from Amravati University, MH, India, in 2006, and the M.Tech degree in Telecommunications Systems and Ph.D. degree in wireless networks from IIT Kharagpur, WB, India, in 2009 and 2016, respectively.

He worked as a senior research fellow from Aug. 2009 to March 2015 and a research assistant from Jan. 2016 to Aug. 2017 at the G. S. Sanyal School of Telecommunications (G.S.S.S.T), IIT Kharagpur. From Sept. 2017 to June 2019, he worked as a postdoctoral research associate in Prof. Harpreet Dhillon's research group at the Department of Electrical and Computer Engineering at Virginia Tech, USA. His research interests primarily focus on the modeling and analysis of wireless networks using the tools of stochastic geometry.

Harpreet S. Dhillon (S'11–M'13–SM'19) received the B.Tech. degree in electronics and communication engineering from IIT Guwahati in 2008, the M.S. degree in electrical engineering from Virginia Tech in 2010, and the Ph.D. degree in electrical engineering from the University of Texas at Austin in 2013.

After serving as a Viterbi Postdoctoral Fellow at the University of Southern California for a year, he joined Virginia Tech in 2014, where he is currently an Associate Professor of electrical and computer engineering and the Elizabeth and James E. Turner Jr. '56 Faculty Fellow. His research interests include communication theory, wireless networks, stochastic geometry, and machine learning. He is a Clarivate Analytics Highly Cited Researcher and has coauthored five best paper award recipients including the 2014 IEEE Leonard G. Abraham Prize, the 2015 IEEE ComSoc Young Author Best Paper Award, and the 2016 IEEE Heinrich Hertz Award. He was named the 2017 Outstanding New Assistant Professor, the 2018 Steven O. Lane Junior Faculty Fellow, and the 2018 College of Engineering Faculty Fellow by Virginia Tech. His other academic honors include the 2008 Agilent Engineering and Technology Award, the UT Austin MCD Fellowship, and the 2013 UT Austin Wireless Networking and Communications Group leadership award. He currently serves as a Senior Editor for the IEEE WIRELESS COMMUNICATIONS LETTERS and an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING.