

Cell-Free Massive MIMO with Finite Fronthaul Capacity: A Stochastic Geometry Perspective

Priyabrata Parida and Harpreet S. Dhillon

Abstract—In this work, we analyze the downlink performance of a cell-free massive multiple-input-multiple-output system with finite capacity fronthaul links between the centralized baseband unit and the access point (APs). Conditioned on the user and AP locations, we first derive an achievable rate for a randomly selected user in the network that captures the effect of finite fronthaul capacity as a compression error. From this expression, we establish that for the traditional cell-free architecture where each AP serves all the users in the network, the achievable rate becomes zero as the network size grows. Hence, to have a meaningful analysis, for the traditional architecture, we model the user and AP locations as two independent binomial point processes over a finite region and provide an accurate theoretical result to determine the user rate coverage. In contrast, for an asymptotically large network, we consider a user-centric architecture where each user in the network is served by a specified number of nearest APs that limits the fronthaul load. For this architecture, we model the AP and user locations as two independent Poisson point processes (PPPs). Since the rate expression is a function of the number of users served by an AP, we statistically characterize the load in terms of the number of users per AP. As the exact derivation of the probability mass function of the load is intractable, we first present the exact expressions for the first two moments of the load. Next, we approximate the load as a negative binomial random variable through the moment matching method. Using the load results along with appropriate distance distributions of a PPP, we present an accurate theoretical expression for the rate coverage of the typical user. From the analyses, we conclude that for the traditional architecture when the AP transmit power is relatively high, a more collocated antenna deployment is preferred. Further, for the user-centric architecture, the energy efficiency of the system is a concave function of the number of antennas per AP.

Index Terms—Cell-free massive MIMO, stochastic geometry, limited fronthaul, binomial point process, Poisson point process, AB random geometric graph.

I. INTRODUCTION

Cooperative cellular networks, where a set of multiple base stations (BSs)/ access points (APs) simultaneously serve a set of users, have been the subject of much investigation throughout the last decade. The latest incarnation of such networks is the cell-free massive multiple-input-multiple-output (mMIMO) systems that harness the benefits of network densification by extending the concept of cellular mMIMO to a distributed implementation [3], [4]. In cell-free mMIMO

networks, the APs perform a limited set of signal processing operations such as precoding/filtering using the local channel state information (CSI) while most of the baseband processing operations are carried out at the centralized baseband units (BBUs). The communication between the APs and BBUs is done through finite capacity fronthaul links. One of the direct consequences of having finite fronthaul links is that the compression/quantization error gets introduced into the system, which affects the user performance. Hence, analyzing the network-wide performance of cell-free mMIMO with finite fronthaul capacity is an important requirement for the successful integration of this technology to the fifth generation (5G) and beyond networks. In this work, our goal is to model and analyze such a system using tools from stochastic geometry and provide useful system design guidelines.

A. Related works

We first discuss key prior works that focus on devising compression algorithms while taking into account the limited fronthaul capacity for other variants of cooperative cellular networks such as coordinated multipoint (CoMP) and cloud radio access networks (C-RAN). In [5], [6], the authors provide information-theoretic insights regarding the capacity of a backhaul-constrained distributed MIMO system. In other notable works, authors use optimization-based frameworks to devise compression algorithms that efficiently utilize the fronthaul capacity constraints while maximizing a certain performance metric (e.g., sum-rate) (cf. [7], [8]). A comprehensive overview of such works can be found in [9], [10]. While the insights obtained from these works are useful, they are not all directly applicable to a cell-free mMIMO system owing to its unique aspects such as beamforming based on local imperfect CSI at the APs as well as the time division duplex (TDD) mode of operation. A consequence of these aspects is a completely different user signal-to-interference-plus noise (SINR) expression compared to the system-level analyses of CoMP and C-RAN. This motivated a separate set of system-level analyses [11]–[19] for cell-free mMIMO with finite fronthaul capacity as briefly outlined below.

In [11], the authors analyze the uplink performance using Bussgang decomposition and studies the effect of the number of quantization bits on the uplink outage probability. In [12], the authors extend the framework of [11] and compare the uplink performance of the scheme where both the quantized version of the received signal and quantized channel estimates are available at the BBU to the scheme where the BBU has the quantized weighted signal from each AP. In

Authors are with Wireless@VT, Dept. of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA (Email: {pparida, hdhillon}@vt.edu). The support of the US NSF (Grant ECCS-1731711) is gratefully acknowledged. This paper was presented in part at the IEEE Vehicular Technology Conference (VTC-Fall), 2018 [1]. A preliminary version of this work appears as a chapter in P. Parida's Ph.D. Thesis [2].

In addition, an uplink max-min power allocation algorithm and an AP-user assignment scheme to reduce the fronthaul load are also proposed. In [15], the authors compare the uplink performances of the following three cases: perfect fronthaul links, when the quantized version of the estimated channel and signal are available at the BBU, and when only quantized weighted signal is available at the BBU. The uplink energy efficiency of a cell-free network with finite fronthaul capacity is analyzed in [13]. In [14], the authors propose precoding and pilot allocation schemes for a mmWave cell-free network. In [16], the authors study the performance of a cell-free network with hardware impairments and limited fronthaul capacity. The uplink and downlink performance of fronthaul constrained cell-free network with low resolutions ADCs is studied in [17]. In [19], the authors compare the downlink performance between local conjugate beamforming and centralized zero-forcing precoding. Note that most of these works focus on *traditional* cell-free architecture where all the APs serve each user in the network. Since the user performance degrades with quantization/compression error, which depends on the number of users (load) per AP, each AP should ideally serve only a subset of users in the network. A network-centric approach that achieves this goal is proposed in [12], [15]. However, from the perspective of scalability and distributed implementation, a *user-centric* architecture is preferred where a user selects its set of serving APs [18], [20]–[27]. To the best of our knowledge, the downlink performance of the user-centric cell-free architecture with finite fronthaul capacity has not been studied in the literature yet. Given its importance, one of our objectives is to bridge this gap in the literature.

From the perspective of system-level analysis, a complementary approach to simulations-based studies is theoretical analyses using tools from stochastic geometry. To this end, there has been a lot of work that analyzes the performance of cooperative cellular networks, such as CoMP and C-RAN (cf. [28]–[33]). However, as mentioned earlier, the system architecture and key practical constraints of cell-free mMIMO, such as imperfect CSI, local beamforming, finite fronthaul capacity, result in a different SINR expression compared to the aforementioned works. Hence, the analyses developed in these works cannot be directly extended to the performance analysis of cell-free mMIMO system. Finally, the performance of cell-free architecture has been analyzed using tools from the stochastic geometry in the following works [23], [34]–[38]. In [34], the authors have focused on characterizing the channel hardening phenomenon in cell-free mMIMO and have presented simulation-based results for system performance metrics such as average user rate. In [23], the authors have used stochastic geometry to study the benefits of the proposed “on-the-fly” pilot assignment scheme in reducing the effect of pilot contamination and have compared the cell-free mMIMO system performance with a cellular massive MIMO system. In [35], the authors have presented the coverage probability result for a traditional cell-free mMIMO with perfect fronthaul assumption. Further, the deterministic SINR equivalence result presented in this work may not be extended to the user-centric cell-free mMIMO where the typical user is served by a few nearest APs. In [36], authors have used stochastic geometry

to study the performance of a cell-free mMIMO system with wireless information and power transfer capabilities. Authors of [37] have considered a cell-free mMIMO network with high mobility users that leads to channel aging effect and degradation in achievable user rate both in the uplink and downlink. Using tools from stochastic geometry, authors have carried out frame length optimization that maximizes downlink and uplink data rates under different transmission protocols. For all these works, the performance analysis is carried out assuming perfect fronthaul capacity, which is one of the key practical assumptions of our work. Although in [38], authors have analyzed the performance of a mmWave cell-free network considering the effect of finite fronthaul, the analysis is carried out for the uplink. Moreover, the traditional architecture is considered where each AP serves each user in the network.

B. Contributions

With the background on aforementioned prior works, our contributions are as follows:

1) *System modeling*: In this work, we consider the downlink of a cell-free mMIMO system with finite capacity fronthaul links. To capture the effect of finite fronthaul, we consider a point-to-point compression scheme between an AP and the BBU. Further, we focus on both the traditional cell-free mMIMO architecture, where each AP serves each user in the network, and a variant of the user-centric cell-free architecture. Since the compression error is a function of the number of users, the traditional cell-free network has to be of finite size in order to limit the effect of compression error. Hence, for this architecture, we model the AP and user locations as two independent binomial point processes (BPPs) that is in line with most of the works in the cell-free literature where fixed numbers of APs and users are considered. On the other hand, for the user-centric architecture, we are not restricted to consider a small network size. Therefore, we model the AP and user locations as two independent homogeneous Poisson point processes (PPPs) on $\mathbb{R}^2$ and assume that each user is served by a specified number of its nearest APs. We restrict our attention to conjugate beamforming. Conditioned on the AP and user locations, we derive an achievable rate expression for a randomly selected user that captures the effect of finite fronthaul capacity in both the architectures.

2) *Load characterization of user-centric architecture*: Due to the dependence of compression error on the number of users served by an AP, the statistics of the load in terms of the number of users is important for the system-level analysis. While this number is fixed for the traditional architecture, the load is a function of user and AP densities as well as the number of APs that serve a given user in the user-centric architecture. Hence, for the latter, we first determine the load distribution for the set of tagged APs that serve the typical user. Since an exact determination of the probability mass function (PMF) of the number of users associated with each tagged AP is intractable, we derive the first two moments of the load and then approximate load for each of the tagged APs as a negative binomial random variable through the moment matching method. This result is later used to derive the rate

coverage of the typical user in the user-centric architecture. Further, we use a similar methodology to derive the load result for the typical AP in the network. This result is useful in network dimensioning, especially determining the desired capacity of the fronthaul link between the typical AP and the BBU to satisfy a certain signal to compression noise ratio (SCNR). It is worth mentioning that this result has a direct equivalence to the degree distribution in an *AB random geometric graph*.

3) Performance analysis of the traditional architecture:

Using the above results, we first derive the downlink user rate coverage result for the traditional architecture. Leveraging the relevant distance distributions for a BPP, we provide an approximate expression to analytically evaluate the rate coverage averaged over the AP and user locations. From our analyses, we infer that the average system sum-rate is a strictly quasi-concave function of the number of users, and the optimal number of users to achieve the maximum system sum-rate increases with increasing fronthaul capacity. Further, in contrast to the established notion that fully distributed MIMO is superior to the collocated MIMO, our results suggest that at higher per AP transmit power, a less distributed form of cell-free mMIMO is better, i.e., for an equal number of antennas in the system, it is better to deploy a fewer APs with more antennas per AP.

4) Performance analysis of the user-centric architecture::

Using the load distribution result of the typical AP, we highlight the interplay between key system parameters such as the fronthaul capacity, the SCNR, and the number of serving APs. Further, exploiting the statistical properties associated with a PPP along with a few subtle approximations, we derive the rate coverage result for the typical user. A key ingredient of this derivation is the load distribution results for the tagged APs. Our analysis suggests that the energy efficiency of the system is a concave function of the number of antennas per AP.

II. SYSTEM MODEL

We limit our attention to the downlink of a cell-free mMIMO system. The sets of AP and user locations are given by Φ_r and Φ_u , respectively. To capture the spatial randomness in the AP and user locations, we model Φ_r and Φ_u by appropriate point processes. The corresponding discussions on the point processes are relegated to the following sections as it is not necessary for the results derived in this section. We assume that each AP has N_a antennas. The distance between a user at $\mathbf{u}_k \in \Phi_u$ and an AP at $\mathbf{r}_m \in \Phi_r$ is denoted by d_{mk} . All the APs are connected to a BBU through a fronthaul network, where the capacity of each link is C_f bits/s/Hz. As mentioned earlier, in case of the traditional cell-free architecture, all the APs serve all the users in the network. In contrast, in case of the user-centric network architecture, we consider that each user is served by its nearest N_s APs. Both the architectures are illustrated in Fig. 1.

A. Compression at the BBU

Due to limited fronthaul capacity, the BBU employs a lossy compression scheme to forward user symbols to the APs.

Let an AP at $\mathbf{r}_o$ serves a set of K_o users $\Phi_{uo} \subseteq \Phi_u$. Note that in the case of traditional architecture, $\Phi_{uo} = \Phi_u$. Let $\mathbf{q}_o = [q_{1o}, q_{2o}, \dots, q_{K_o}]^T$ be the signal vector consisting of the symbols to be transmitted to the users in Φ_{uo} . We consider that $\mathbf{q}_o$ is a circularly symmetric complex Gaussian random vector and $\mathbf{q}_o \sim \mathcal{CN}(\mathbf{0}_{K_o}, \rho_{q_o} \mathbf{I}_{K_o})$, where $\rho_{q_o} = \mathbb{E}[|q_{1o}|^2] = \mathbb{E}[|q_{2o}|^2] = \dots = \mathbb{E}[|q_{K_o}|^2]$, $\mathbf{0}_{K_o}$ denotes a $K_o \times 1$ all zero vector, and $\mathbf{I}_{K_o}$ denotes a $K_o \times K_o$ identity matrix. Using a lossy compression scheme, the BBU transmits $\hat{\mathbf{q}}_o = [\hat{q}_{1o}, \hat{q}_{2o}, \dots, \hat{q}_{K_o}]^T$ over the fronthaul links to the AP. Similar to [5], we consider $\hat{\mathbf{q}}_o = \mathbf{q}_o + \tilde{\mathbf{q}}_o$, where $\tilde{\mathbf{q}}_o \sim \mathcal{CN}(\mathbf{0}_{K_o}, \rho_{\tilde{q}_o} \mathbf{I}_{K_o})$ is the compression error vector and $\rho_{\tilde{q}_o} = \mathbb{E}[|\tilde{q}_{1o}|^2] = \mathbb{E}[|\tilde{q}_{2o}|^2] = \dots = \mathbb{E}[|\tilde{q}_{K_o}|^2]$. Further, we assume that $\mathbf{q}_o$ and $\tilde{\mathbf{q}}_o$ are uncorrelated. Since both are Gaussian random vectors, they are independent as well. From the above exposition, it is clear that $\hat{\mathbf{q}} \sim \mathcal{CN}(\mathbf{0}_{K_o}, (\rho_{\tilde{q}_o} + \rho_{q_o}) \mathbf{I}_{K_o})$. If $\mathbb{E}[|\hat{q}_{k_o}|^2]$ is the same for all $k = 1, 2, \dots, K_o$, then both $\rho_{\tilde{q}_o}, \rho_{q_o}$ depend on the fronthaul capacity C_f , as discussed in the following lemma.

Lemma 1. For a fronthaul capacity C_f and number of users K_o served by the typical AP, $\rho_{q_o} = (1 - 2^{-C_f/K_o}) \mathbb{E}[|\hat{q}_{k_o}|^2]$ and $\rho_{\tilde{q}_o} = 2^{-C_f/K_o} \mathbb{E}[|\hat{q}_{k_o}|^2]$.

Proof: The amount of information that can be transmitted from the BBU to each AP is upper bounded by the fronthaul capacity C_f . Hence, we write $I(\hat{\mathbf{q}}_o; \mathbf{q}_o) \leq C_f$

$$\begin{aligned} &\Rightarrow h(\hat{\mathbf{q}}_o) - h(\hat{\mathbf{q}}_o | \mathbf{q}_o) \leq C_f \\ &\Rightarrow \sum_{i=1}^{K_o} h(\hat{q}_{i_o}) - \sum_{i=1}^{K_o} h(\hat{q}_{i_o} | q_{i_o}) \leq C_f \\ &\Rightarrow \log_2(\pi e(\rho_{q_o} + \rho_{\tilde{q}_o})) - \log_2(\pi e \rho_{\tilde{q}_o}) \leq \frac{C_f}{K_o}, \end{aligned}$$

where $I(x; y)$ denotes the mutual information between two random variables x and y , $h(x)$ denotes the differential entropy of a random variable x , and the last step follows from the fact that $\hat{q}_{i_o}$ s and $\tilde{q}_{i_o}$ s are complex Gaussian random variables. Ideally, the BBU would like to transmit the maximum information. Hence, we get

$$\log_2 \left(1 + \frac{\rho_{q_o}}{\rho_{\tilde{q}_o}} \right) = \frac{C_f}{K_o} \Rightarrow \frac{\rho_{q_o}}{\rho_{\tilde{q}_o}} = 2^{C_f/K_o} - 1.$$

The expression in the lemma follows directly using the fact that $\rho_{q_o} + \rho_{\tilde{q}_o} = \mathbb{E}[|\hat{q}_{k_o}|^2]$. If we consider that $\mathbb{E}[|\hat{q}_{k_o}|^2] = 1$, then $\rho_{q_o} = (1 - 2^{-C_f/K_o})$ and $\rho_{\tilde{q}_o} = 2^{-C_f/K_o}$. ■

Remark 1. The SCNR, defined as $\frac{\rho_{q_o}}{\rho_{\tilde{q}_o}} = 2^{C_f/K_o} - 1$, is a decreasing function of the number of users served by the AP. While in the case of the traditional cell-free mMIMO, the SCNR can only be improved by increasing C_f , in the case of user-centric architecture, SCNR can also be improved by limiting the maximum number of users that should be scheduled by the typical AP. Hence, for a given C_f and target SCNR threshold T_s , the maximum number of scheduled users $K_{\max}$ should satisfy $K_{\max} \log_2(1 + T_s) \leq C_f$.

B. Uplink channel estimation

Let $\mathbf{g}_{mk} = \sqrt{\beta_{mk}} \mathbf{h}_{mk}$ be the channel gain between the AP at $\mathbf{r}_m$ and the user at $\mathbf{u}_k$, where β_{mk} captures the large-scale

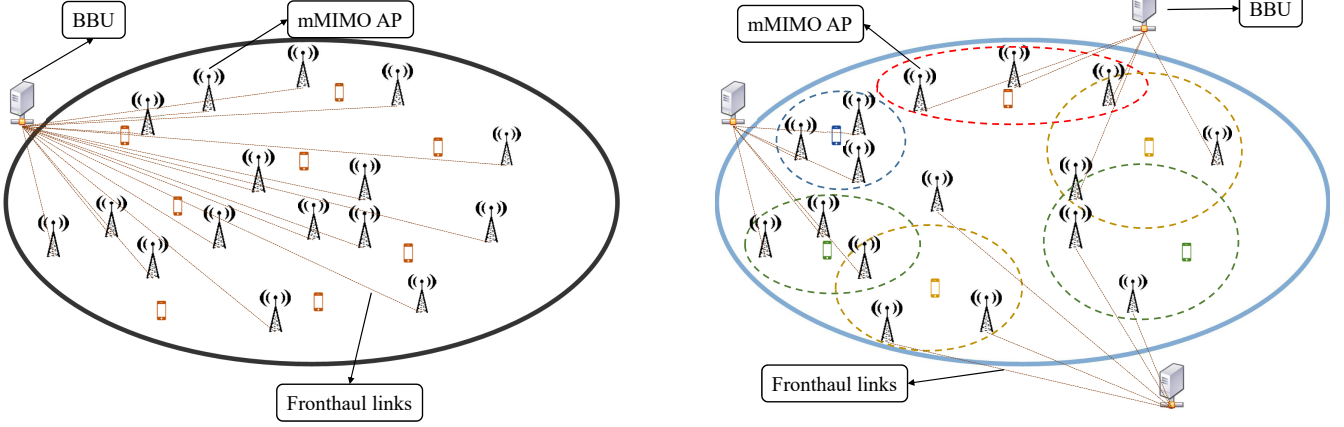


Fig. 1: Representative diagrams: (left) in the traditional architecture, each AP serves all the user in the network, and (right) in the user-centric architecture, each user is served by its nearest three APs as marked by dotted circles.

channel gain and $\mathbf{h}_{mk} \sim \mathcal{CN}(\mathbf{0}_{N_a}, \mathbf{I}_{N_a})$ captures the small-scale channel fluctuations. We consider that β_{mk} is only due to the distance dependent path loss, i.e., $\beta_{mk} = l(d_{mk})^{-1}$, where $l(\cdot)$ is a non-decreasing path loss function presented in Section VI.

In order to obtain the channel estimates, we consider that each user uses a pilot from a set of P orthogonal pilot sequences of τ_p symbol duration, which is assumed to be less than the length of the coherence block τ_c . Further, the transmit signal-to-noise ratio (SNR) of each symbol in a pilot is ρ_p . Since we assume that these P sequences are orthogonal to each other, $\tau_p \geq P$ and $\psi_i^H \psi_j = \mathbf{1}(i = j)$, where $\mathbf{1}(\cdot)$ denotes the indicator function. Let the pilot used by the user at $\mathbf{u}_k$ be $\psi(k)$. During the pilot transmission phase, the received signal matrix $\mathbf{R}_o \in \mathbb{C}^{N_a \times \tau_p}$ at the typical AP is

$$\mathbf{R}_o = \sqrt{\tau_p \rho_p} \sum_{\mathbf{u}_k \in \Phi_u} \mathbf{g}_{ok} \psi(k)^T + \mathbf{W}_o,$$

where each element of $\mathbf{W}_o$ is $\mathcal{CN}(0, 1)$. Let $\hat{\mathbf{g}}_{ok}$ be the estimated channel vector at the AP $\mathbf{r}_o$ for the user $\mathbf{u}_k \in \Phi_u$ that is obtained after performing minimum-mean-squared-error (MMSE) channel estimation. Further, $\tilde{\mathbf{g}}_{ok}$ be the estimation error vector. Using the properties of MMSE estimation [4], we write $\hat{\mathbf{g}}_{ok} \sim \mathcal{CN}(\mathbf{0}_{N_a}, \gamma_{ok} \mathbf{I}_{N_a})$ and $\tilde{\mathbf{g}}_{ok} \sim \mathcal{CN}(\mathbf{0}_{N_a}, (\beta_{ok} - \gamma_{ok}) \mathbf{I}_{N_a})$, where

$$\gamma_{ok} = \frac{\tau_p \rho_p \beta_{ok}^2}{1 + \tau_p \rho_p \sum_{\mathbf{u}_j \in \Phi_u} \psi(k)^H \psi(j) \beta_{oj}}. \quad (1)$$

C. Downlink data transmission

In this work, we consider that each AP employs conjugate beamforming based on the local CSI. Hence, the precoded symbol transmitted by the AP at $\mathbf{r}_o$ is given as

$$\mathbf{x}_o = \sum_{\mathbf{u}_i \in \Psi_{uo}} \sqrt{\rho_d \eta_{oi}} \frac{\hat{\mathbf{g}}_{oi}^*}{\sqrt{\mathbb{E}[\|\hat{\mathbf{g}}_{oi}\|^2]}} \hat{q}_{i_o} = \sum_{\mathbf{u}_i \in \Psi_{uo}} \sqrt{\rho_d \eta_{oi}} \mathbf{w}_{oi} \hat{q}_{i_o},$$

where ρ_d is the downlink transmit SNR, η_{oi} is normalization coefficient used by the typical AP for the user at $\mathbf{u}_i$ to satisfy the average power constraint $\text{Tr}(\mathbb{E}[\mathbf{x}_o \mathbf{x}_o^H]) \leq \rho_d$, and $\Psi_{uo} \subseteq$

Φ_{uo} is the set of scheduled users associated with the AP at $\mathbf{r}_o$ such that $|\Psi_{uo}| \leq K_{\max}$ for the user-centric architecture. Note that for the traditional architecture, $\Psi_{uo} = \Phi_{uo} = \Phi_u$ and $K_{\max} = K_o$. We observe that by setting $\eta_{mk} = 1/K_{\max}$ and $\rho_{\hat{q}_o} = \mathbb{E}[|\hat{q}_{i_o}|^2] = 1$, the above constraint is satisfied. More sophisticated power allocation algorithms, such as max-min power allocation, can of course be considered. However, the equal power allocation scheme has its own advantages of lower complexity and admitting a distributed implementation. Besides, this scheme provides a degree of tractability in the coverage analysis as will be evident in the sequel.

D. An achievable rate for a randomly selected user

Now, we present an achievable rate for a randomly selected user in the network that is applicable for both types of architectures. Consider that a randomly selected user is located at $\mathbf{u}_o$ and is served by the set of APs $\Phi_{ro} \subseteq \Phi_r$. The received signal at this user is given as

$$\begin{aligned} y_o^{\text{dl}} &= \sum_{\mathbf{r}_l \in \Phi_{ro}} \mathbf{g}_{lo}^T \mathbf{x}_l + \sum_{\mathbf{r}_j \in \Phi_{ro}^C} \mathbf{g}_{jo}^T \mathbf{x}_j + \mathbf{n}_o \\ &= \sum_{\mathbf{r}_l \in \Phi_{ro}} \sqrt{\rho_d \eta_{lo}} \frac{\mathbf{g}_{lo} \hat{\mathbf{g}}_{lo}^*}{\sqrt{N_a \gamma_{lo}}} \hat{q}_{l_o} + \sum_{\mathbf{r}_l \in \Phi_{ro}^C} \sum_{\mathbf{u}_i \in \Psi_{ul}} \frac{\sqrt{\rho_d \eta_{li}} \mathbf{g}_{li}^T \hat{\mathbf{g}}_{li}^*}{\sqrt{N_a \gamma_{li}}} \hat{q}_{li} \\ &\quad + \sum_{\mathbf{r}_l \in \Phi_{ro}^C} \sum_{\mathbf{u}_i \in \Phi_{ul}} \frac{\sqrt{\rho_d \eta_{li}} \mathbf{g}_{li}^T \hat{\mathbf{g}}_{li}^*}{\sqrt{N_a \gamma_{li}}} \hat{q}_{li} + n_o, \end{aligned} \quad (2)$$

where $\Phi_{ro}^C = \Phi_r \setminus \Phi_{ro}$ and $\tilde{\Psi}_{ul} = \Psi_{ul} \setminus \mathbf{u}_o$. In the following lemma, we provide an expression for an achievable rate (a lower bound on capacity).

Lemma 2. Conditioned on Φ_r and Φ_u , an achievable rate of the typical user at $\mathbf{u}_o$ is given as

$$\text{SE}_o = \xi_{\text{td}} \left(1 - \frac{\tau_p}{\tau_c} \right) \log_2 (1 + \text{SINR}_o) \quad \text{bits/s/Hz}, \quad (3)$$

where SINR_o is given in (4) at the top of the next page. Further, in (4), $\mathcal{P}_o$ is the set of users that use the same pilot sequence as the typical user $\mathbf{u}_o$ and k_l is the number of users served by

$$\text{SINR}_o = \frac{\rho_d N_a \left(\sum_{\mathbf{r}_l \in \Phi_{ro}} \sqrt{\frac{\gamma_{lo}(1-2^{-C_t/k_l})}{K_{\max}}} \right)^2}{\rho_d N_a \sum_{\mathbf{r}_l \in \Phi_{ro}} \gamma_{lo} \frac{2^{-C_t/k_l}}{K_{\max}} + \rho_d \sum_{\mathbf{r}_l \in \Phi_r} \beta_{lo} + \rho_d N_a \sum_{\mathbf{u}_i \in \{\mathcal{P}_o \setminus \mathbf{u}_o\}} \left(\sum_{\mathbf{r}_l \in \Phi_{ri}} \sqrt{\frac{\gamma_{lo}}{K_{\max}}} \right)^2 + 1} \quad (4)$$

the AP at $\mathbf{r}_l$, ξ_{tdd} is the fraction of coherence block dedicated for downlink transmission.

Proof: Please refer to Appendix A. ■

Note that the above SINR expression contains the interference due to pilot contamination that originates due to reuse of pilot sequences in the network. The presence of pilot contamination makes the derivation of the rate coverage expression intractable. Hence, in favor of tractability, we ignore the effect of pilot contamination in subsequent analyses.

III. RATE COVERAGE FOR TRADITIONAL CELL-FREE mMIMO

In this section, we derive the rate coverage result for the traditional cell-free mMIMO system where each AP serves all the users in the network. If we consider an infinite network on $\mathbb{R}^2$, then as per the result of Lemma 1, the SCNR $\rightarrow 0$ as $K_o \rightarrow \infty$ and subsequently $\text{SINR}_o \rightarrow 0$ as given in Lemma 2. Hence, for a meaningful analysis of the traditional architecture, we need to consider a finite network, e.g., a shopping mall. Therefore, we assume the system is limited to $\mathcal{B}_{R_s}(\mathbf{o})$, a finite circular region of radius R_s centered at $\mathbf{o}$, where the set of APs $\Phi_r = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_M\}$ are randomly and uniformly distributed. Further, $\Phi_u = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{K_o}\}$ contains the set of user locations that are uniformly and randomly distributed in $\mathcal{B}_{R_s}(\mathbf{o})$ and are independent of AP locations. Note that by construction, Φ_r and Φ_u form two independent BPPs. Alternatively, one can consider modeling Φ_r and Φ_u as two independent PPPs over a finite region. Since conditioned on the number of points, a PPP is equivalent to a BPP, it is straightforward to extend the results of this section to the finite PPP-based model. In that case, the statistical performance metrics need further averaging over the number of APs M and the number of users K_o that follow the Poisson distribution. With this fundamental understanding, we consider the BPP-based modeling of Φ_r and Φ_u for further discussion. The BPP-based assumption is also consistent with most works related to the traditional cell-free mMIMO literature that consider a fixed number of users and APs in the network. As assumed in the cell-free mMIMO literature, we consider that $MN_a \gg K_o$. Further, we assume that the coherence block is sufficiently long to ensure that $\tau_p \geq K_o^{-1}$. As a consequence, pilots are not reused in the network, thereby eliminating the effect of pilot contamination. Under these assumptions, using the result

of Lemma 2, the achievable rate of a user at $\mathbf{u}_o$ is given as $\text{SE}_{o,\text{fin}} = \hat{\xi}_{\text{tdd}} \log_2(1 + \text{SINR}_{o,\text{fin}})$, where

$$\text{SINR}_{o,\text{fin}} = \frac{\rho_d \frac{N_a}{K_o} (1 - 2^{-C_t/K_o}) \left(\sum_{m=1}^M \sqrt{\gamma_{mo}} \right)^2}{\rho_d \frac{N_a}{K_o} 2^{-C_t/K_o} \sum_{m=1}^M \gamma_{mo} + \rho_d \sum_{m=1}^M \beta_{mo} + 1} \quad (5)$$

is the SINR of the user at $\mathbf{u}_o$ in this finite network and $\hat{\xi}_{\text{tdd}} = \xi_{\text{tdd}} \left(1 - \frac{\tau_p}{\tau_c}\right)$

Our goal is to determine the rate coverage $R_{c,\text{fin}}(T_r) = \mathbb{P}[\hat{\xi}_{\text{tdd}} \log_2(1 + \text{SINR}_{o,\text{fin}}) > T_r]$ for a randomly selected user that requires averaging over the distances of the APs from the user. Hence, now we present a few relevant distance distributions for a BPP.

A. Relevant distance distributions in a BPP

Let R_o be the distance of the user at $\mathbf{u}_o$ from the center of the circle $\mathcal{B}_{R_s}(\mathbf{o})$. Since $\mathbf{u}_o$ is uniformly and randomly distributed in $\mathcal{B}_{R_s}(\mathbf{o})$, the cumulative distribution function (CDF) and probability density function (PDF) of R_o is given as

$$F_{R_o}(r) = r^2/R_s^2, \text{ and } f_{R_o}(r) = 2r/R_s^2, \quad 0 \leq r \leq R_s. \quad (6)$$

Next, we present the distance distribution between $\mathbf{u}_o$ to a randomly distributed AP in $\mathcal{B}_{R_s}(\mathbf{o})$.

Lemma 3. Conditioned on the distance R_o , the CDF of the distance between the user at $\mathbf{u}_o$ and the AP at $\mathbf{r}_m$ is given as

$$F_{D_{mo}}(d|r_o) = \left(\frac{d^2 \left(\theta^* - \frac{\sin(2\theta^*)}{2} \right)}{\pi R_s^2} + \frac{\left(\phi^* - \frac{\sin(2\phi^*)}{2} \right)}{\pi} \right) \times \mathbf{1}_{R_s - r_o \leq d \leq R_s + r_o} + \frac{d^2}{R_s^2} \mathbf{1}_{0 \leq d < R_s - r_o},$$

and corresponding PDF is given as

$$f_{D_{mo}}(d|r_o) = \frac{2d}{R_s^2} \mathbf{1}_{0 \leq d < R_s - r_o} + \frac{2d\theta^*}{\pi R_s^2} \mathbf{1}_{R_s - r_o \leq d \leq R_s + r_o}$$

where $\theta^* = \arccos\left(\frac{d^2 + r_o^2 - R_s^2}{2r_o d}\right)$, $\phi^* = \arccos\left(\frac{R_s^2 + r_o^2 - d^2}{2r_o R_s}\right)$, and $\mathbf{1}_{(\cdot)}$ is the indicator function.

Proof: We provide a proof sketch for this lemma. Please refer to [39, Lemma 1] for the detailed proof. Without loss of generality, consider that $\mathbf{u}_o = (r_o, 0)$. Then, conditioned on $\mathbf{u}_o$ (equivalently r_o), a uniformly distributed point in $\mathcal{B}_{R_s}(\mathbf{o})$ can lie either in the circle $\mathcal{B}_{R_s - r_o}(\mathbf{u}_o)$ or in the region $\mathcal{B}_{R_s}(\mathbf{o}) \setminus \mathcal{B}_{R_s - r_o}(\mathbf{u}_o)$. In the CDF expression of the lemma, both these conditions are captured by the indicator function and corresponding conditional CDFs are presented. The expression

¹ If we assume a coherence block that spans 200 kHz - 500 kHz in frequency and 2 ms in time, then the signaling dimension is between 400 - 1000 symbols. For the considered finite system model, by reserving 20% of the symbols, we can serve 80 - 200 users in the network without pilot contamination.

for the PDF is obtained by taking the derivative of the CDF with respect to d along with some algebraic manipulation. ■

Now, using the results from order statistics, we present the conditional distance distribution between $\mathbf{u}_o$ and its nearest AP.

Lemma 4. *Conditioned on the distance R_o , the CDF of the distance D_{oo} between $\mathbf{u}_o$ and its nearest AP is given as $F_{D_{oo}}(d_{oo}|r_o) = \mathbb{P}[D_{oo} \leq d_{oo}|r_o] = 1 - (1 - F_{D_{mo}}(d_{oo}|r_o))^M$, and the corresponding PDF is given as $f_{D_{oo}}(d_{oo}|r_o) = M f_{D_{mo}}(d_{oo}|r_o)(1 - F_{D_{mo}}(d_{oo}|r_o))^{M-1}$, where $f_{D_{mo}}, F_{D_{mo}}$ were presented in Lemma 3.*

Note that conditioned on the distance D_{oo} , rest of the APs in $\mathcal{B}_{R_s}(\mathbf{o})$ are uniformly and randomly located in $\mathcal{B}_{R_s}(\mathbf{o}) \setminus \mathcal{B}_{d_{oo}}(\mathbf{u}_o)$, where d_{oo} is a realization of D_{oo} . In the following lemma, we present the distribution of the distance between a randomly located AP in the above region and $\mathbf{u}_o$.

Lemma 5. *Conditioned D_{oo} and R_o , the PDF of the distance $\hat{D}_{mo}$ between a randomly selected AP in $\mathcal{B}_{R_s}(\mathbf{o}) \setminus \mathcal{B}_{d_{oo}}(\mathbf{u}_o)$ and $\mathbf{u}_o$ is given as*

$$f_{\hat{D}_{mo}}(d|d_{oo}, r_o) = \frac{f_{D_{mo}}(d|r_o)}{1 - F_{D_{mo}}(d_{oo}|r_o)}, \quad d_{oo} \leq d \leq r_o + R_s.$$

Proof: We provide a proof sketch for this lemma. For the detailed proof, please refer to [39, Lemma 3]. Conditioned on D_{oo} , rest of the APs are uniformly distributed in $\mathcal{B}_{R_s}(\mathbf{o}) \setminus \mathcal{B}_{d_{oo}}(\mathbf{u}_o)$. Hence, the distribution of the distance $\hat{D}_{mo}$ follows the lower truncated distribution of D_{mo} , which is captured in the above expression. ■

Next, using the above distance distributions, we derive the rate coverage expression.

B. Approximate evaluation of average achievable user rate

The exact evaluation of rate coverage $R_{c,\text{fin}}$ is challenging as it requires an $(M+1)$ -fold integration to average it over the locations of all the M APs and the user at $\mathbf{u}_o$. Notice that the $\text{SINR}_{o,\text{fin}}$ in (5) has the following terms:

$$I_1 = \sum_{m=1}^M \sqrt{\gamma_{mo}}, \quad I_2 = \sum_{m=1}^M \gamma_{mo}, \quad I_3 = \sum_{m=1}^M \beta_{mo}. \quad (7)$$

Since there is no pilot contamination, $\gamma_{mk}(d_{mk}) = \frac{\tau_p \rho_p l(d_{mk})^{-2}}{1 + \tau_p \rho_p l(d_{mk})^{-1}}$. Further, γ_{mk} is a decreasing function of d_{mk} . Due to path loss, these terms are dominated by contributions from a few nearest APs. Hence, we approximate I_1, I_2 , and I_3 as the sum of exact contribution from the nearest AP and the mean contribution from the rest of the APs conditioned on the distance d_{oo} between $\mathbf{u}_o$ and its nearest AP. Hence, we write

$$\begin{aligned} I_1(d_{oo}, r_o) &\approx \sqrt{\gamma_{oo}} + \mathbb{E} \left[\sum_{m=1, m \neq o}^M \sqrt{\gamma_{mo}} \middle| d_{oo}, r_o \right], \\ I_2(d_{oo}, r_o) &\approx \gamma_{oo} + \mathbb{E} \left[\sum_{m=1, m \neq o}^M \gamma_{mo} \middle| d_{oo}, r_o \right], \\ I_3(d_{oo}, r_o) &\approx \beta_{oo} + \mathbb{E} \left[\sum_{m=1, m \neq o}^M \beta_{mo} \middle| d_{oo}, r_o \right]. \end{aligned} \quad (8)$$

It is worth mentioning that this approach based on dominant interferers has been extensively used in the stochastic geometry literature (cf. [40], [41]). Note that conditioned on D_{oo} , distances between $\mathbf{u}_o$ and rest of the APs in the network are i.i.d [39, Lemma 3]. Hence, using Campbell's theorem, (8) can be written as

$$\begin{aligned} \hat{I}_1(d_{oo}, r_o) &= \sqrt{\gamma_{oo}} + (M-1) \\ &\quad \times \int_{r=d_{oo}}^{r_o+R_s} \frac{\sqrt{\tau_p \rho_p} l(r)^{-1}}{\sqrt{1 + \tau_p \rho_p} l(r)^{-1}} f_{\hat{D}_{mo}}(r|d_{oo}, r_o) dr, \\ \hat{I}_2(d_{oo}, r_o) &= \gamma_{oo} + (M-1) \\ &\quad \times \int_{r=d_{oo}}^{r_o+R_s} \frac{\tau_p \rho_p l(r)^{-2}}{1 + \tau_p \rho_p l(r)^{-1}} f_{\hat{D}_{mo}}(r|d_{oo}, r_o) dr, \\ \hat{I}_3(d_{oo}, r_o) &= \beta_{oo} + (M-1) \\ &\quad \times \int_{r=d_{oo}}^{r_o+R_s} l(r)^{-1} f_{\hat{D}_{mo}}(r|d_{oo}, r_o) dr. \end{aligned} \quad (9)$$

With the above approximation, in the next Proposition, we present an expression to evaluate the rate coverage of the typical user in this finite cell-free mMIMO network.

Proposition 1. *For a given threshold T_r , the rate coverage of a randomly selected user in the network is given as*

$$R_{c,\text{fin}}(T_r) = \int_{r_o=0}^{R_s} \int_{d_{oo}=0}^{R_s} \mathbf{1} \left(\text{SINR}_o^{\text{APX}}(d_{oo}, r_o) > 2^{T_r/\xi_{\text{tad}}} - 1 \right) \times f_{D_{oo}}(d_{oo}|r_o) f_{R_o}(r_o) dd_{oo} dr_o,$$

where $\text{SINR}_o^{\text{APX}}(d_{oo}, r_o) =$

$$\frac{\rho_d \frac{N}{K_o} (1 - 2^{-C_t/K_o}) (\hat{I}_1(d_{oo}, r_o))^2}{\rho_d \frac{N}{K_o} (\hat{I}_2(d_{oo}, r_o)) 2^{-C_t/K_o} + \rho_d \hat{I}_3(d_{oo}, r_o) + 1}, \quad (10)$$

and the PDFs of D_{oo} and R_o are presented in Lemma 4 and (6), respectively.

Proof: This result follows by first replacing different terms in the $\text{SINR}_{o,\text{fin}}$ by their approximations given in (9) to obtain $\text{SINR}_o^{\text{APX}}(d_{oo}, r_o)$. In the next step, we decondition over D_{oo} and R_o to obtain $R_{c,\text{fin}}(T_r)$. ■

Next we present the metric to study the energy efficiency of the traditional cell-free mMIMO system that use the result of Proposition 1.

C. Energy efficiency of the network

For the finite traditional cell-free network, we define energy efficiency as the ratio of average network throughput to the average total power consumption at the APs and their associate fronthaul. We ignore the signal processing-related power consumption at the central baseband unit as it is likely to be small compared to the aforementioned AP-related power consumption in the network. Further, since we focus on the downlink, we do not take into account the uplink pilot transmission power of the users in our power consumption model. Hence, for the downlink, the total power consumption for the AP at $\mathbf{r}_l \in \Phi_r$ can be modeled as [42]

$$P_{l,\text{tot}} = \Delta_a \frac{|\Phi_u|}{K_o} P_d + P_{l,\text{fix}}$$

$$+ \sum_{\mathbf{u}_k \in \Phi_{\mathbf{u}}} P_{\text{fh},v} B_{\text{W}} \hat{\xi}_{\text{td}} \log_2(1 + \text{SINR}_{o,\text{fin}}), \quad (11)$$

where the first term on the right hand side corresponds to total downlink transmission power and $\Delta_{\text{a}} > 1$ captures the power amplifier inefficiency. Note that for the finite network the APs use all their transmission power P_{a} , which is related to the transmit SNR $\rho_{\text{a}} = P_{\text{a}}/P_{\text{n}}$, where P_{n} is the total thermal noise power. The second term in (11) corresponds to the both circuit power consumption of the AP and the traffic independent power consumption of the fronthaul link associated with the AP and can be expressed as $P_{l,\text{fix}} = N_{\text{a}}\kappa_{\text{rf}} + P_{\text{fh},f}$, where κ_{rf} is the RF chain power consumption that can be assumed to be the same for all the antennas. Further, $P_{\text{fh},f}$ corresponds to the fixed power consumption of the fronthaul link. The third term in (11) captures the traffic dependent power consumption of the fronthaul link, where B_{W} is the system bandwidth and $P_{\text{fh},v}$ is a scaling constant. Now, for the downlink, the average total power consumption of the network is given as $P_{\text{tot}} = \mathbb{E}[\sum_{\mathbf{r}_l \in \Phi_{\text{r}}} P_{l,\text{tot}}]$. Putting all terms together, for the finite traditional cell-free network, the average energy efficiency is given in (12) at the top of the next page. In Section VI, we evaluate the performance of the traditional cell-free mMIMO in terms of the metrics derived in this section.

Next, we focus on the user-centric cell-free mMIMO. In this case, we model Φ_{r} and Φ_{u} as two independent homogeneous PPPs on $\mathbb{R}^2$. The densities of Φ_{r} and Φ_{u} are λ_{r} and λ_{u} , respectively. Further, each user is served by its nearest N_{s} APs. Observe that the SINR expression of Lemma 2 is a function of the number of users served by each AP in the network. Owing to the spatial randomness of both user and AP locations, the number of users served by each AP is a random variable. Therefore, to derive the rate coverage expression, we need the statistical properties of the load associated with an AP that are presented in the next section.

IV. LOAD CHARACTERIZATION IN USER-CENTRIC ARCHITECTURE

Before proceeding further, we provide the distinction between the *typical AP* and the set of *tagged APs* in the network. The typical AP is by definition a randomly selected AP in Φ_{r} . On the other hand the set of tagged APs is the set of serving APs of the typical user, which is selected randomly from Φ_{u} . This random selection of the typical user makes it more likely to be served by APs that have larger service regions. This effect is reminiscent of the *waiting time* paradox in queuing systems and the difference between 0-cell and the typical cell in a Poisson-Voronoi tessellation [43], [44]. Since we are focusing on the performance analysis of the typical user, we need the load distribution result associated with the set of tagged APs. On the other hand, from a network dimensioning perspective, such as provisioning of fronthaul capacity, we also need the statistical information on the load associated with the typical AP. In the next subsection, we first derive the load distribution results associated with the set of tagged APs.

A. The load of a tagged AP

The statistical metric that we are interested in is the PMF of the load. The exact derivation of PMF is intractable. Hence,

we first derive the exact result for the first two moments of the number of users for a tagged AP.

1) *Determination of the first two moments:* Since Φ_{u} is a homogeneous PPP, it is translation invariant. Hence, we assume that the typical user is located at the origin $\mathbf{o}$. It is worth mentioning that the loads associated with each of the tagged APs are not identical. Hence, we need to present a generic result that is a function of the serving AP rank in terms of the distance from the typical user. Next, we derive the first two moments of the load for the N -th nearest tagged AP.

Lemma 6. *The first moment of the number of users (excluding the typical user) served by the N -th nearest AP to the typical user at the origin is given as $\mathbb{E}[K_N] =$*

$$2\pi\lambda_{\text{r}}\lambda_{\text{u}} \int_{r_o=0}^{\infty} dr_o \int_{d_x=0}^{\infty} dd_x \int_{v_x=0}^{2\pi} dv_x h_{\text{tag},m_1}(r_o, d_x, v_x, N) dx r_o,$$

where $h_{\text{tag},m_1}(r_o, d_x, v_x, N)$ is given in (13) at the top of the next page. In the above expression $r_x(r_o, d_x, v_x) = \sqrt{r_o^2 + d_x^2 - 2r_o d_x \cos(v_x)}$ and $\text{AoI}_2(r_o, d_x, v_x)$ is the area of intersection of two circles is given as

$$\begin{aligned} \text{AoI}_2(r_o, d_x, v_x) = & r_x^2 \left(u(r_o, d_x, v_x) - \frac{\sin(2u(r_o, d_x, v_x))}{2} \right) \\ & + r_o^2 \left(v_x - \frac{\sin(2v_x)}{2} \right), \end{aligned} \quad (14)$$

where

$$u(r_1, r_2, v) = \arccos \left(\frac{r_2 - r_1 \cos(v)}{\sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos(v)}} \right). \quad (15)$$

The corresponding second moment is given as $\mathbb{E}[K_N^2] =$

$$\begin{aligned} 2\pi\lambda_{\text{r}}\lambda_{\text{u}}^2 \int_{r_o=0}^{\infty} \int_{d_x=0}^{\infty} \int_{d_y=0}^{\infty} \int_{v_x=0}^{2\pi} \int_{v_y=0}^{2\pi} \\ \times h_{\text{tag},m_2}(r_o, d_x, d_y, v_x, v_y, N) dv_y dv_x dd_y dx dd_x r_o dr_o, \end{aligned}$$

where $h_{\text{tag},m_2}(r_o, d_x, d_y, v_x, v_y, N)$ is given by (26) in Appendix B.

Proof: Please refer to Appendix B. ■

Remark 2. *The load on a tagged AP depends on its distance from the typical user at the origin. Using the results of the above lemma, we conclude that $\mathbb{E}[K_1] > \mathbb{E}[K_2] > \mathbb{E}[K_3] > \dots$, and $\mathbb{E}[|K_1|^2] > \mathbb{E}[|K_2|^2] > \mathbb{E}[|K_3|^2] > \dots$*

2) *Approximation of the load PMF:* Using the first two moments, we approximate the load as a negative binomial random variable. The PMF of this variable with parameters r and p is

$$\mathbb{P}[K_N = k] = \frac{\Gamma(k+r)}{k!\Gamma(r)} p^r (1-p)^k,$$

where $\mathbb{E}[K_N] = (1-p)r/p$ and $\mathbb{E}[K_N^2] = (1-p)r(1+(1-p)r)/p^2$. Using the results of Lemma 6, we solve the aforementioned two equations to obtain the values of r and p . The intuition behind the consideration of negative binomial

$$\mathbb{E}E_{\text{fin}} = \frac{B_w |\Phi_u| \hat{\xi}_{\text{td}} \mathbb{E}[\log_2(1 + \text{SINR}_{o,\text{fin}})]}{\Delta_a |\Phi_r| P_d + |\Phi_r| (N_a \xi_l + P_{\text{fh},s}) + P_{\text{fh},v} B_w |\Phi_u| |\Phi_r| \hat{\xi}_{\text{td}} \mathbb{E}[\log_2(1 + \text{SINR}_{o,\text{fin}})]}. \quad (12)$$

$$h_{\text{tag},m_1}(r_o, d_x, v_x, N) = \sum_{n=0}^{N-1} \text{Pos}_{\text{PMF}}(N - n - 1, \lambda_r(\pi r_o^2 - \text{AoI}_2(r_o, d_x, v_x))) \\ \text{Pos}_{\text{PMF}}(n, \lambda_r \text{AoI}_2(r_o, d_x, v_x)) \text{Pos}_{\text{CMF}}(N_s - n - 1, \lambda_r(\pi r_x^2 - \text{AoI}_2(r_o, d_x, v_x))). \quad (13)$$

stems from the following fact: if we consider that each user is served by its nearest AP, then the serving region of each AP is its own Poisson-Voronoi cell. The area of this cell is well approximated as a gamma random variable [45]. Now, conditioned on the area of the cell, the number of users that fall in this area follows Poisson distribution. Hence, once we decondition over the serving area, we get a Poisson-gamma mixture distribution for the number of users served by a tagged AP. Since negative binomial distribution is a consequence of the Poisson-gamma mixture distribution, this justifies our choice to approximate the load with this distribution. This approximation will be validated in Sec. VI.

B. The load of the typical AP

In this section, we derive the approximate PMF of the number of users served by the typical AP in the network. Similar to the previous case, since exact characterization of the PMF is intractable, we first derive the exact expression for the first two moments of the load for the typical AP. Next, using moment-matching method, we approximate the PMF as a negative binomial PMF. The derivation of the first two moments now becomes the special case of the tagged AP result. In the following lemma, we present the first two moments.

Lemma 7. *The first two moments of the number of users K_o served by the typical AP are*

$$\mathbb{E}[K_o] = N_s \frac{\lambda_u}{\lambda_r}, \\ \mathbb{E}[K_o^2] = 2\pi \lambda_u^2 \int_{r_x=0}^{\infty} \int_{r_y=0}^{\infty} \int_{u=0}^{2\pi} h_{\text{typ},m_2}(r_1, r_2, u) dr_2 dr_1 du \\ + N_s \frac{\lambda_u}{\lambda_r},$$

where $h_{\text{typ},m_2}(r_x, r_y, u)$ is given in (16) at the top of the next page. In the above equation, $v_{xy} = \arccos\left(\frac{r_x - r_y \cos(u)}{\sqrt{r_x^2 + r_y^2 - 2r_x r_y \cos(u)}}\right)$, and AoI_2 is given in (14).

Proof: Please refer to Appendix C. ■

Similar to the tagged AP case, we approximate the load of the typical AP as a negative binomial random variable.

Remark 3. We observe that $\mathbb{E}[K_o^2] \approx \mathbb{E}[K_o] + (1 + (0.2802)^{N_s}) \mathbb{E}^2[K_o]$. This result is known for $N_s = 1$ [46].

Remark 4. The load distribution result of the typical AP also characterizes the degree distribution in an AB random

geometric graph (AB-RGG). An AB-RGG is a bipartite random graph between two sets of vertices A and B , where a point in A is connected to a few points in B based on certain distance criteria [47], [48]. In our case, $A = \Phi_r$, $B = \Phi_u$ and an edge exists if $\mathbf{x} \in \Phi_u$ is served by $\mathbf{y} \in \Phi_r$. It is worth mentioning that a simulation-based approximation result for the degree distribution of this type of AB-RGG is recently studied in [49].

V. RATE COVERAGE FOR USER-CENTRIC CELL-FREE MMIMO

The achievable rate result derived in Lemma 2 is directly applicable to the user-centric cell-free architecture. Recall that Φ_r and Φ_u are two independent homogeneous PPP. Further, we consider that the typical user at the origin and its set of serving APs $\Phi_{r,o}$ consists of the nearest N_s APs. Determining the distribution of the rate or SINR of the typical user is intractable as each term in (4) depends on a set of common distances. However, a degree of tractability can be achieved for theoretical analysis by assuming that the network is operating in a regime where pilot contamination is negligible.

Remark 5. A system with a pilot allocation scheme that ensures that each AP serves only one user per pilot is likely to operate in the regime where the above assumption holds, especially, for the set of serving (dominant) APs for the typical user. This assignment scheme is realizable in a low mobility scenario, where the coherence block is sufficiently large. For example, if we consider a coherence bandwidth of 200 kHz and a coherence time of 2 ms, then the TDD coherence block has 400 symbols. Let us assume that each AP is served by the nearest $N_s = 5$ APs. In such a scenario, the probability that the set of tagged APs, which serve the typical user, collectively serve more than 30 users is less than 0.002. Hence, the above criterion is met by reserving around 30 symbols, which is less than 10% of the coherence block, for pilot estimation.

With this assumption, we present the rate coverage for the typical user.

Proposition 2. The rate coverage $R_{c,\text{inf}}(T_r)$ for the typical user for a target threshold T_r is given in (17) at the top of the next page. In the expression, $d_{1o} \leq d_{2o} \leq \dots \leq d_{N_s o}$, where d_{io} is the distance between the typical user and its i -th closest AP. The load of the closest AP is K_1 and the mean load of the i -th closest AP is $\bar{K}_i$, which is given as $\bar{K}_i = 1 + \sum_{k_i=0}^{\infty} \min\{k_i, K_{\text{max}}\} \mathbb{P}[K_i = k_i]$.

Proof: Please refer to Appendix E. ■

$$h_{\text{typ},m2}(r_x, r_y, u) = \sum_{l=0}^{N_s-1} \left[\text{Pos}_{\text{PMF}}(l, \lambda_r \text{AoI}_2(r_x, r_y, v_{xy})) \text{Pos}_{\text{CMF}}(N_s - l - 1, \lambda_r(\pi r_x^2 - \text{AoI}_2(r_x, r_y, v_{xy}))) \right. \\ \left. \times \text{Pos}_{\text{CMF}}(N_s - l - 1, \lambda_r(\pi r_y^2 - \text{AoI}_2(r_x, r_y, v_{xy}))) \right]. \quad (16)$$

$$\mathbb{R}_{c,\text{inf}}(T_r) = \mathbb{P}[\text{SE}_o > T_r] = \int_{d_{N_s o}=0}^{\infty} \int_{d_{N_s-1 o}=0}^{d_{N_s o}} \dots \int_{d_{1 o}=0}^{d_{N_s o}} \sum_{k_1=0}^{\infty} \mathbf{1} \left(2\pi \lambda_r \int_{d_{o N_s}}^{\infty} l(r) r dr \leq h_{\text{cov}}(k_1, d_{1 o}, d_{2 o}, \dots, d_{N_s o}) \right) \\ \times \mathbb{P}[K_1 = k_1] \prod_{i=1}^{N_s-1} \frac{2d_{i o}}{d_{N_s o}} f_{D_{N_s o}}(d_{N_s o}) dd_{1 o} \dots dd_{N_s o}, \quad (17)$$

where

$$h_{\text{cov}}(k_1, d_{1 o}, d_{2 o}, \dots, d_{N_s o}) = \frac{N_a}{(2^{T_r/\hat{\xi}_{\text{tdd}}} - 1)K_{\text{max}}} \left(\sqrt{\gamma_{1 o}(1 - 2^{-\frac{C_f}{\min\{k_1+1, K_{\text{max}}\}}})} + \sum_{l=2}^{N_s} \sqrt{\gamma_{l o}(1 - 2^{-C_f/\bar{K}_l})} \right) \\ - \frac{N_a}{K_{\text{max}}} \left(\gamma_{1 o} 2^{-\frac{C_f}{\min\{k_1+1, K_{\text{max}}\}}} + \sum_{l=2}^{N_s} \gamma_{l o} 2^{-C_f/\bar{K}_l} \right) - \sum_{l=1}^{N_s} l(d_{l o}) - \frac{1}{\rho_d}. \quad (18)$$

A. Energy efficiency of the user-centric cell-free mMIMO

For the user-centric network, the downlink energy efficiency can be defined as the ratio of average throughput per unit area to the average power consumption per unit area [50]. The average throughput per unit area is equal to $\lambda_u B_w \hat{\xi}_{\text{tdd}} \mathbb{E}[\log_2(1 + \text{SINR}_{0_k})]$, where $\mathbb{E}[\log_2(1 + \text{SINR}_{0_k})]$ is the average user spectral efficiency of the typical user that is evaluated using Proposition 2. Further, the average power consumption at the typical AP can be modeled as [42]

$$P_{l,\text{tot}} = \frac{\Delta_a}{K_{\text{max}}} P_d \frac{\lambda_u N_s}{\lambda_r} + (N_a \kappa_{\text{rf}} + P_{\text{fh},f}) \\ + P_{\text{fh},v} B_w \frac{\lambda_u N_s}{\lambda_r} \hat{\xi}_{\text{tdd}} \mathbb{E}[\log_2(1 + \text{SINR}_{0_k})],$$

where the first term captures the average transmission power of the AP while taking into account the power amplifier inefficiency, the second term captures the traffic independent power consumption related to the RF chains and fronthaul link, and the third captures the traffic dependent power consumption of the fronthaul link. The average power consumption per unit area is $P_{\text{tot}} = \lambda_r P_{l,\text{tot}}$. Combining all terms together, for the user-centric cell-free mMIMO, the energy efficiency is given in (19) at the top of the next page. In the following section, using the derived results of this section we provide useful system design guidelines for the user-centric cell-free mMIMO network.

VI. RESULTS AND DISCUSSION

In this section, we provide useful system design insights from our analysis as well as validate our theoretical rate coverage results through extensive Monte Carlo simulations. Unless otherwise stated, solid lines and markers represent analytical and simulation results, respectively.

A. Performance of traditional cell-free mMIMO

First, we validate the theoretical results derived for the traditional architecture in Sec. III. We have considered a finite circular region of radius $R_s = 500$ m. We have used 3GPP Urban Micro-cell path-loss function [51] which is given as

$$10 \log 10(l(r)) = 30.5 + 36.7 \log_{10}(\max(r, 1)). \quad (20)$$

We consider a maximum transmit SNR $\rho_d = 122$ dB and transmit pilot SNR $\rho_p = 112$ dB. The system bandwidth B_w is 20 MHz and the noise power is set to -92 dBm over this bandwidth. Hence, the maximum AP transmit power $P_d = 1$ watt. Let the TDD coherence block consists of $\tau_c = 200$ symbols, which corresponds to a coherence bandwidth of 200 kHz and a coherence time of 1 ms. We assume that the users are assigned orthogonal pilot sequences, equivalently the number of users in this finite system $K \leq \tau_p$. The fraction ξ_{tdd} of TDD resource allocated for the downlink is 0.8. The choice of other system parameters are indicated at necessary places. Using the rate coverage result of Proposition 1, the average user rate is expressed as

$$\bar{\text{SE}}_{o,\text{fin}} = \int_{t=0}^{\infty} t \mathbb{R}_{c,\text{fin}}(t) dt \quad \text{bits/s/Hz},$$

and corresponding average system sum-rate is $K \bar{\text{SE}}_{o,\text{fin}}$ bits/s/Hz.

1) *The effect of fronthaul capacity:* In Fig. 2 (left), we have presented the rate coverage of the system for $K = 20$ users in the system. As expected, the rate coverage improves with increasing SCNR threshold T_s as it directly corresponds to a higher fronthaul capacity C_f . Further, in Fig. 2 (right), the average system sum-rate is presented as a function of the number of users K for different fronthaul capacities. As evident from the figure, the average system sum-rate is quasi-concave function of the number of users. Further, for a given

$$EE_{\text{inf}} = \frac{\lambda_u B_w \hat{\xi}_{\text{td}} \mathbb{E}[\log_2(1 + \text{SINR}_{0_k})]}{\frac{\Delta_a}{K_{\text{max}}} P_d N_s \lambda_u + \lambda_r (N_a \kappa_{\text{rf}} + P_{\text{fh},f}) + P_{\text{fh},v} B_w \lambda_u N_s \hat{\xi}_{\text{td}} \mathbb{E}[\log_2(1 + \text{SINR}_{0_k})]}. \quad (19)$$

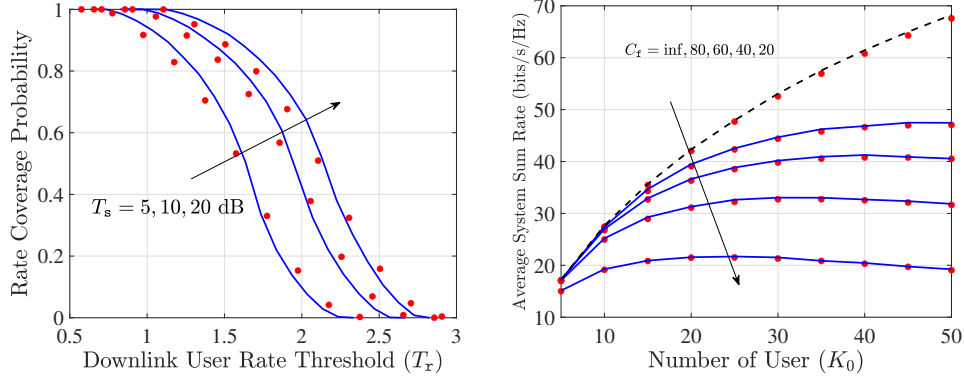


Fig. 2: The effect of fronthaul capacity on system performance. System parameters: $M = 600$, $N_a = 40$, $\tau_p = 20$ (left), $\tau_p = 50$ (right), $K_o = 20$, $\rho_p = 112$ dB, $\rho_a = 122$ dB.

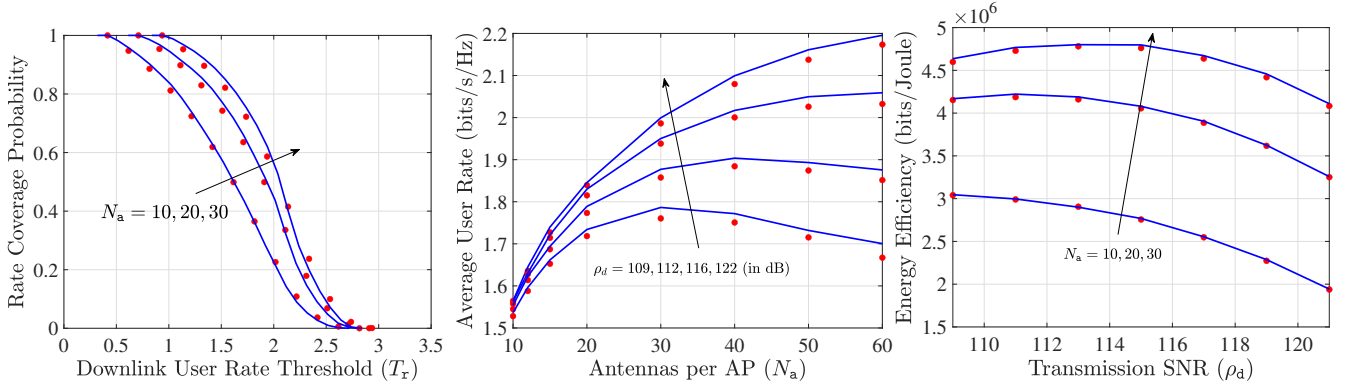


Fig. 3: The effect of number of antennas per AP on rate coverage (left) and average user rate (center). We have considered $MN_a = 600$, $\tau_p = 40$, $K = 20$, $\rho_p = 112$ dB, $\text{SCNR} = 20$ dB. (Right) The effect of transmit SNR on the energy efficiency. Parameters: $MN_a = 600$, $K_o = 20$, $\tau_p = 40$, $\rho_p = 112$ dB, $\text{SCNR} = 20$ dB, $B_w = 20$ MHz.

number of APs, the optimum number of users that should be multiplexed to maximize the average system rate increases with the increasing fronthaul capacity.

2) *Distributed vs. collocated*: In Fig. 3 (left), we present the rate coverage of the system for different number of antennas at the AP while keeping the total number of antennas in the system fixed, i.e., $MN_a = 600$. We observe that the centralized architecture performs better than a distributed architecture under conjugate beamforming and equal power allocation. In Fig. 3 (center), we present the average user rate for different number of antennas at each AP while keeping the total number of antennas in the service region fixed. We consider the SCNR threshold $T_s = 20$ dB. We observe that for high transmit SNR ρ_a , average user rate increases as we move towards a more collocated setup. On the other hand, with low ρ_a , the average user SE is a concave function of the number of antennas per AP. This behavior is in contrast to the conventional MIMO results where a distributed implementation is always preferable. The justification to this counter-intuitive trend can be explained by the fact that due to conjugate beamforming, we get a self-interference term from all the APs as evident

from the SINR expression in (5). Hence, with a distributed implementation, the desired signal power from the nearest AP increases and so does the self interference term. Therefore, a more collocated set up is preferable for high ρ_a .

3) *Energy efficiency*: In Fig. 3 (right), we present the energy efficiency of the traditional cell-free mMIMO with respect to transmit SNR. As evident from the figure, a distributed network is less energy efficient. Further, the energy efficiency is a quasi-concave function of transmit SNR.

B. Performance of user-centric cell-free mMIMO

Now, we verify the theoretical results derived in Secs. IV and V through extensive Monte Carlo simulations. Further, we provide a few useful network dimensioning guidelines. For the simulations, we consider the system radius to be 2000 m. The path loss is the same as (20). The choice of other system parameters are indicated at appropriate places.

1) *The load distribution result*: In Fig. 4, we validate the approximate theoretical PMFs derived in Secs. IV-A and IV-B, respectively, by Monte Carlo simulations. In both the cases,

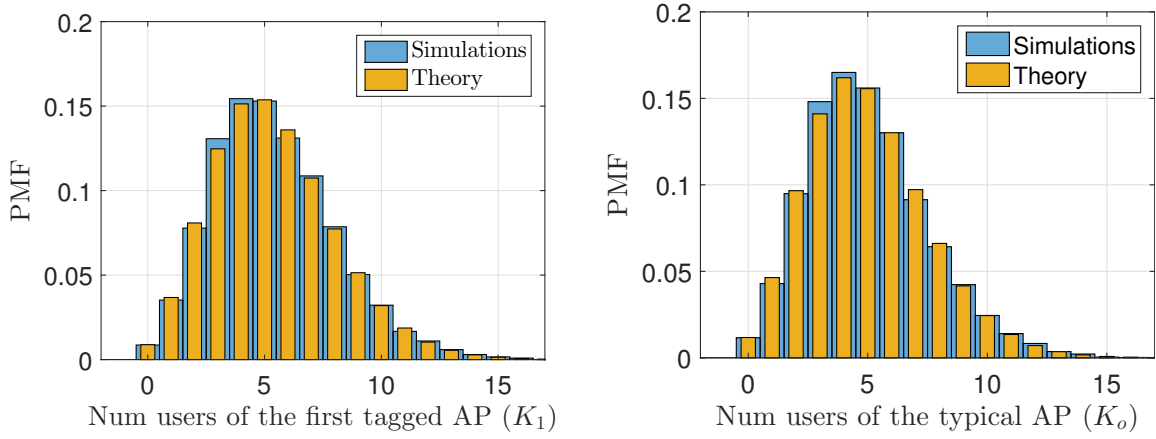


Fig. 4: The PMF of the load of the nearest tagged AP (left) and the typical AP (right) in the network. $\lambda_r = \lambda_u = 10^{-4}$, $N_s = 5$.

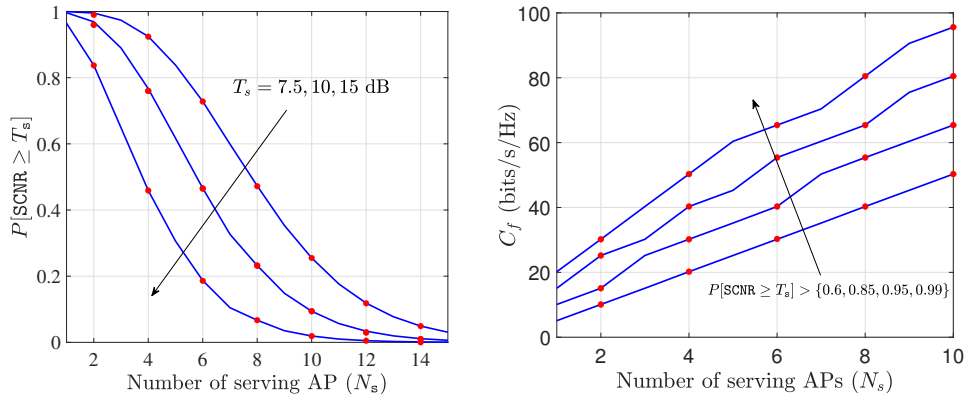


Fig. 5: (Left) The number of serving APs to ensure a certain minimum SCNR for different T_s . Other parameters: $\lambda_u = \lambda_r = 10^{-4}$, $C_f = 20$ bits/s/Hz. (Right) The required fronthaul capacity as a function of N_s to ensure $\mathbb{P}[\text{SCNR} \geq T_s]$ is above a certain threshold. Other parameters: $\lambda_u = \lambda_r = 10^{-4}$, $T_s = 15$ dB.

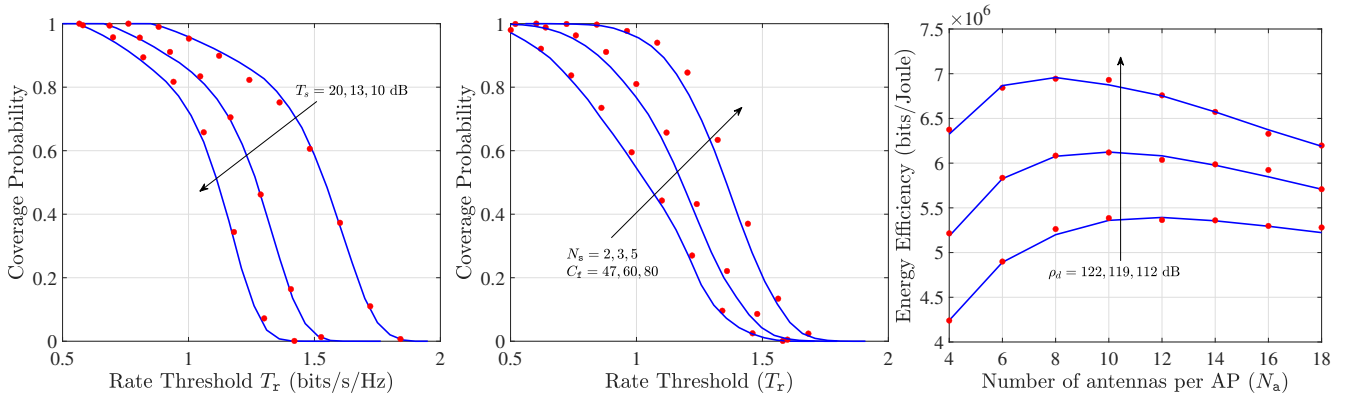


Fig. 6: (Left) The rate coverage of the typical user for different T_s and a fixed $C_f = 80$ bits/s/Hz. Other parameters: $\lambda_u = \lambda_r = 10^{-4}$, $N_a = 20$, $N_s = 5$, $\tau_p = 40$, $T_c = 200$; $\rho_p = 112$ dB, $\rho_d = 122$ dB. (Center) The rate coverage for different number of serving APs (N_s). (Right) Energy efficiency of the user-centric cell-free system as a function of N_a . Other parameters: $\lambda_u = \lambda_r = 10^{-4}$, $N_s = 5$, $C_f = 80$ bits/s/Hz, $B_w = 20$ MHz.

theoretical and simulation results are remarkably close. In Fig. 5 (left), using the typical AP load distribution result, for a given fronthaul capacity C_f and SCNR threshold T_s , we present the probability that SCNR is above T_s as a function of number of serving APs N_s . Formally, we study $\mathbb{P}[\text{SCNR} \geq T_s] =$

$$\mathbb{P}[2^{C_f/K_o} - 1 \geq T_s] = \mathbb{P}[K_o \leq C_f / \log_2(1 + T_s)],$$

which is the CDF of the typical AP load. As expected, the more stringent the T_s , the lower the probability of having a SCNR more than T_s for a given number of serving APs. Note that in this result we assume that all the users attached to the typical AP are scheduled on the same resource while using different pilots. In Fig. 5 (right), we present the required fronthaul capacity as a function of number of serving APs. As

expected, we observe a linear growth in C_f with increasing N_s . However, the rate of growth depends on how strict the SCNR constraint is.

2) *The rate coverage and energy efficiency results:* In Fig. 6 (left and center), we present the rate coverage result for the user-centric architecture. The results suggest that the simulations and the theory are in close agreement. As expected, with reducing T_s coverage probability improves as the APs are allowed to transmit to a potentially lesser number of users thereby reducing the network interference. Further, increasing N_s also improves the coverage probability. In Fig. 6 (right), we present the system energy efficiency result highlighting its concave nature with respect to the number of antennas per AP.

VII. CONCLUSION

In this work, we modeled and analyzed a cell-free mMIMO network with finite fronthaul capacity using tools from stochastic geometry. We considered two different architectures of cell-free mMIMO, namely, the traditional architecture where each AP serves all the users in the network, and the user-centric architecture where the typical user is served by a few nearest APs. For the traditional architecture, we derived the user rate coverage result using the relevant statistics of BPPs. For the user-centric architecture, we statistically characterized the load for the typical AP as well as the set of tagged APs that serve the typical user in the network. Further, using the statistical properties of PPPs, we presented the rate coverage result for the typical user in the network. From our analyses, we conclude that for the traditional architecture, when the AP transmit power is relatively high, a more collocated implementation of cell-free mMIMO is preferred over the fully distributed implementation. Further, for the user-centric architecture, there exists an optimal number of antennas per AP that maximizes the system energy efficiency. Promising future directions include consideration of more practical fading scenarios such as Rician fading. Moreover, the analysis presented in this work can be extended to consider other beamforming schemes such as full-pilot zero-forcing. The system analysis in the presence of pilot contamination is also another promising direction that can lead to useful guidelines for pilot allocation.

APPENDIX

A. Proof of Lemma 2

The proof of this lemma is based on a lower bound that is well known in the mMIMO literature (cf. [52, Lemma 2], [53]). From (2), we can express y_o^{d1} as given in (21) at the top of the next page, where we assume that the user $\mathbf{u}_o$ has the knowledge of the average channel statistics with respect to its serving APs and $\mathcal{P}_o$ contains the user locations that use the same pilot sequence as $\mathbf{u}_o$. In (21), it can be shown that the desired signal is uncorrelated to the rest of the signals. To derive an achievable rate (lower bound on the capacity), we replace the uncorrelated signals by independent Gaussian random variables [53]. The variance of each of the Gaussian variables is equal to the corresponding variance of the undesired signal. Hence, the SINR corresponding to this achievable

rate is given as $\text{SINR}_o = \mathbb{E}[|T_1|^2] / (\sum_{i=2}^7 \mathbb{E}[|T_i|^2] + 1)$. In this case, note that $\mathbb{E}[T_i] = 0$ for all i . Further,

$$\begin{aligned} \mathbb{E}[|T_1|^2] &= \rho_d N_a \left(\sum_{\mathbf{r}_l \in \Phi_{r_o}} \sqrt{\frac{\gamma_{lo}(1 - 2^{-C_f/k_l})}{K_{\max}}} \right)^2, \\ \mathbb{E}[|T_2|^2] &= \rho_d \sum_{\mathbf{r}_l \in \Phi_{r_o}} \frac{\gamma_{lo}(1 - 2^{-C_f/k_l})}{K_{\max}}, \\ \mathbb{E}[|T_3|^2] &= \rho_d (N_a + 1) \sum_{\mathbf{r}_l \in \Phi_{r_o}} \frac{\gamma_{lo}}{K_{\max}} 2^{-C_f/k_l}, \\ \mathbb{E}[|T_4|^2] &= \rho_d \sum_{\mathbf{r}_l \in \Phi_{r_o}} \frac{(\beta_{lo} - \gamma_{lo})}{K_{\max}}, \\ \mathbb{E}[|T_5|^2] &= \rho_d \sum_{\mathbf{r}_l \in \Phi_{r_o}} \frac{k_l - 1}{K_{\max}} \beta_{lo} \\ &\leq \rho_d \sum_{\mathbf{r}_l \in \Phi_{r_o}} \frac{K_{\max} - 1}{K_{\max}} \beta_{lo} \leq \rho_d \sum_{\mathbf{r}_l \in \Phi_{r_o}} \beta_{lo}, \\ \mathbb{E}[|T_6|^2] &= \rho_d \sum_{\mathbf{r}_l \in \Phi_{r_o}^c} \frac{k_l - 1}{K_{\max}} \beta_{lo} \leq \rho_d \sum_{\mathbf{r}_l \in \Phi_{r_o}^c} \frac{K_{\max} - 1}{K_{\max}} \beta_{lo}, \\ \mathbb{E}[|T_7|^2] &= \sum_{\mathbf{u}_i \in \{\mathcal{P}_o \setminus \mathbf{u}_o\}} \left(\sum_{\mathbf{r}_l \in \Phi_{r_i}} \frac{\rho_d \beta_{lo}}{K_{\max}} + \rho_d N_a \left(\sum_{\mathbf{r}_l \in \Phi_{r_i}} \sqrt{\frac{\gamma_{lo}}{K_{\max}}} \right)^2 \right). \end{aligned}$$

Substituting these values, we obtain the expression presented in the lemma. ■

B. Proof of Lemma 6

The mean load of the N -th closest serving AP to the typical user at $\mathbf{o}$ is given in (22) in the next page, where $r_o = \|\mathbf{r} - \mathbf{o}\|$, $r_x = \|\mathbf{r} - \mathbf{x}\| = \sqrt{r_o^2 + d_x^2 - 2r_o d_x \cos(v_x)}$, $d_x = \|\mathbf{x} - \mathbf{o}\|$ (please refer to Fig. 7 (left)). Conditioned on the location of the AP at $\mathbf{r}$, we expand the inner expectation as given in (23) in the next page, where the second step follows from the fact that $\{\Phi_{\mathbf{r}} \setminus \mathbf{r}\}$ is a homogeneous PPP with density $\lambda_{\mathbf{r}}$ and the regions in each indicator function are non-overlapping. The final result follows in two step: first we decondition over the point $\mathbf{r}$, then we decondition over $\Phi_{\mathbf{u}}$. Formally, $\mathbb{E}[K_N] =$

$$\begin{aligned} &\mathbb{E}_{\Phi_{\mathbf{u}}} \left[\sum_{\mathbf{x} \in \Phi_{\mathbf{u}}} \int_{\mathbf{r} \in \mathbb{R}^2} h_{\text{tag}, \mathbf{m}_1}(\|\mathbf{r}\|, d_x, v_x, N) \lambda_{\mathbf{r}} d\mathbf{r} \right] \\ &= \mathbb{E}_{\Phi_{\mathbf{u}}} \left[\sum_{\mathbf{x} \in \Phi_{\mathbf{u}}} 2\pi \lambda_{\mathbf{r}} \int_{r_o=0}^{\infty} h_{\text{tag}, \mathbf{m}_1}(r_o, d_x, v_x, N) r_o dr_o \right] \\ &= 2\pi \lambda_{\mathbf{u}} \lambda_{\mathbf{r}} \int_{r_o=0}^{\infty} dr_o \int_{d_x=0}^{\infty} dd_x \int_{v_x=0}^{2\pi} dv_x h_{\text{tag}, \mathbf{m}_1}(r_o, d_x, v_x, N) r_o dx, \end{aligned}$$

where the last step follows from the application of Campbell's theorem.

Now, we focus on deriving the second moment of the load $\mathbb{E}[|K_N|^2]$ for the N -th nearest serving AP. Using the similar principle used in (22) for the first moment, we can express $\mathbb{E}[|K_N|^2]$ as given in (24) in the next page. On the right hand side of (24), the second summation term is the mean $\mathbb{E}[K_N]$, which has already been derived above. Hence, we focus on deriving an expression for the first term (please refer to Fig. 7

$$\begin{aligned}
y_o^{\text{dl}} = & \underbrace{\sqrt{\rho_a} \sum_{\mathbf{r}_l \in \Phi_{r_o}} \sqrt{\eta_{lo}} \frac{\mathbb{E}[\|\hat{\mathbf{g}}_{lo}\|^2]}{\sqrt{N_a \gamma_{lo}}} q_{ol}}_{T_1: \text{Desired signal}} + \underbrace{\sqrt{\rho_a} \sum_{\mathbf{r}_l \in \Phi_{r_o}} \sqrt{\eta_{lo}} \frac{(\|\hat{\mathbf{g}}_{lo}\|^2 - \mathbb{E}[\|\hat{\mathbf{g}}_{lo}\|^2])}{\sqrt{N_a \gamma_{lo}}} q_{ol}}_{T_2: \text{Beamforming uncertainty}} + \underbrace{\sum_{\mathbf{r}_l \in \Phi_{r_o}} \frac{\|\hat{\mathbf{g}}_{lo}\|^2}{\sqrt{N_a \gamma_{lo}}} \sqrt{\rho_a \eta_{lo}} \tilde{q}_{ol}}_{T_3: \text{compression error}} \\
& + \underbrace{\sum_{\mathbf{r}_l \in \Phi_{r_o}} \sqrt{\rho_a \eta_{lo}} \frac{\hat{\mathbf{g}}_{lo}^T \hat{\mathbf{g}}_{lo}^*}{\sqrt{N_a \gamma_{lo}}} \hat{q}_{ol}}_{T_4: \text{channel estimation error}} + \underbrace{\sum_{\mathbf{r}_l \in \Phi_{r_o}} \sum_{\mathbf{u}_i \in \Psi_{ul} \setminus \{\mathbf{u}_o\}} \sqrt{\rho_a \eta_{li}} \frac{\mathbf{g}_{lo}^T \hat{\mathbf{g}}_{li}^*}{\sqrt{N_a \gamma_{li}}} \hat{q}_{li}}_{T_5: \text{inter user interference}} + \underbrace{\sum_{\mathbf{r}_l \in \Phi_{r_o}^C} \sum_{\mathbf{u}_i \in \Psi_{ul} \setminus \mathcal{P}_o} \sqrt{\rho_a \eta_{li}} \frac{\mathbf{g}_{lo}^T \hat{\mathbf{g}}_{li}^*}{\sqrt{N_a \gamma_{li}}} \hat{q}_{li}}_{T_6: \text{interfering AP signal}} \\
& + \underbrace{\sum_{\mathbf{u}_i \in \{\mathcal{P}_o \setminus \{\mathbf{u}_o\}\}} \sum_{\mathbf{r}_l \in \Phi_{ri}} \sqrt{\rho_a \eta_{li}} \frac{\mathbf{g}_{lo}^T \hat{\mathbf{g}}_{li}^*}{\sqrt{N_a \gamma_{li}}} \hat{q}_{li}}_{T_7: \text{pilot contamination}} + n_o,
\end{aligned} \tag{21}$$

$$\begin{aligned}
\mathbb{E}[K_N] &= \mathbb{E} \left[\sum_{\mathbf{x} \in \Phi_u} \mathbf{1}(|\Phi_r \cap \mathcal{B}_{r_x}(\mathbf{x})| \leq N_s - 1) \mathbf{1}(|\Phi_r \cap \mathcal{B}_{r_o}(\mathbf{o})| = N - 1) \right] \\
&= \mathbb{E}_{\Phi_u} \left[\sum_{\mathbf{x} \in \Phi_u} \mathbb{E}_{\Phi_r} [\mathbf{1}(|\Phi_r \cap \mathcal{B}_{r_x}(\mathbf{x})| \leq N_s - 1) \mathbf{1}(|\Phi_r \cap \mathcal{B}_{r_o}(\mathbf{o})| = N - 1)] \right],
\end{aligned} \tag{22}$$

$$\begin{aligned}
& \mathbb{E}_{\{\Phi_r \setminus \mathbf{r}\}} \left[\sum_{n=0}^{N-1} \mathbf{1}(|\Phi_r \cap \mathcal{B}_{r_o}(\mathbf{o}) \cap \mathcal{B}_{r_x}(\mathbf{x})| = n) \mathbf{1}(|\Phi_r \cap \{\mathcal{B}_{r_o}(\mathbf{o}) \setminus \mathcal{B}_{r_x}(\mathbf{x})\}| = N - n - 1) \mathbf{1}(|\Phi_r \cap \{\mathcal{B}_{r_x}(\mathbf{x}) \setminus \mathcal{B}_{r_o}(\mathbf{o})\}| \leq N_s - n - 1) \right] \\
&= \sum_{n=0}^{N-1} \frac{(\lambda_r |\mathcal{B}_{r_o}(\mathbf{o}) \cap \mathcal{B}_{r_x}(\mathbf{x})|)^n}{n!} e^{-\lambda_r |\mathcal{B}_{r_o}(\mathbf{o}) \cap \mathcal{B}_{r_x}(\mathbf{x})|} \frac{(\lambda_r |\mathcal{B}_{r_o}(\mathbf{o}) \setminus \mathcal{B}_{r_x}(\mathbf{x})|)^{N-n-1}}{n!} e^{-\lambda_r |\mathcal{B}_{r_o}(\mathbf{o}) \setminus \mathcal{B}_{r_x}(\mathbf{x})|} \\
&\quad \sum_{l=0}^{N_s-n-1} \frac{(\lambda_r |\mathcal{B}_{r_x}(\mathbf{x}) \setminus \mathcal{B}_{r_o}(\mathbf{o})|)^{N_s-n-1-l}}{l!} e^{-\lambda_r |\mathcal{B}_{r_x}(\mathbf{x}) \setminus \mathcal{B}_{r_o}(\mathbf{o})|} \\
&= \sum_{n=0}^{N-1} \text{Pos}_{\text{PMF}}(n, \lambda_r \text{AoI}_2(r_o, d_x, v_x)) \text{Pos}_{\text{PMF}}(N - n - 1, \lambda_r (\pi r_o^2 - \text{AoI}_2(r_o, d_x, v_x))) \\
&\quad \text{Pos}_{\text{CMF}}(N_s - n - 1, \lambda_r (\pi r_x^2 - \text{AoI}_2(r_o, d_x, v_x))) = h_{\text{tag}, \mathbf{m}_1}(r_o = \|\mathbf{r}\|, d_x, v_x, N),
\end{aligned} \tag{23}$$

$$\begin{aligned}
\mathbb{E}[|K_N|^2] &= \mathbb{E} \left[\left| \sum_{\mathbf{x} \in \Phi_u} \mathbf{1}_{|\Phi_r \cap \mathcal{B}_{r_x}(\mathbf{x})| \leq N_s - 1} \mathbf{1}_{|\Phi_r \cap \mathcal{B}_{r_o}(\mathbf{o})| = N - 1} \right|^2 \right] = \mathbb{E} \left[\sum_{\substack{\mathbf{x} \in \Phi_u \\ \mathbf{y} \in \Phi_u}} \mathbf{1}_{|\Phi_r \cap \mathcal{B}_{r_x}(\mathbf{x})| \leq N_s - 1} \mathbf{1}_{|\Phi_r \cap \mathcal{B}_{r_y}(\mathbf{y})| \leq N_s - 1} \mathbf{1}_{|\Phi_r \cap \mathcal{B}_{r_o}(\mathbf{o})| = N - 1} \right] \\
&= \mathbb{E} \left[\sum_{\substack{\mathbf{x} \neq \mathbf{y} \\ \mathbf{x} \in \Phi_u \\ \mathbf{y} \in \Phi_u}} \mathbf{1}_{|\Phi_r \cap \mathcal{B}_{r_x}(\mathbf{x})| \leq N_s - 1} \mathbf{1}_{|\Phi_r \cap \mathcal{B}_{r_y}(\mathbf{y})| \leq N_s - 1} \mathbf{1}_{|\Phi_r \cap \mathcal{B}_{r_o}(\mathbf{o})| = N - 1} \right] + \mathbb{E} \left[\sum_{\mathbf{x} \in \Phi_u} \mathbf{1}_{|\Phi_r \cap \mathcal{B}_{r_x}(\mathbf{x})| \leq N_s - 1} \mathbf{1}_{|\Phi_r \cap \mathcal{B}_{r_o}(\mathbf{o})| = N - 1} \right].
\end{aligned} \tag{24}$$

(center)). Using (15), let us define $u_x = u(r_o, d_x, v_x)$, $u_y = u(r_o, d_y, v_y)$, and

$$u_{xy} = \begin{cases} u_x + u_y, & \{v_x < \pi, v_y > \pi\}, \\ u_x + u_y, & \{v_x > \pi, v_y < \pi\}, \\ |u_x - u_y|, & \text{Otherwise.} \end{cases} \tag{25}$$

Further, let us denote the region of intersection of three circles as $\text{RoI}_{oxy} = \{\mathcal{B}_{r_o}(\mathbf{o}) \cap \mathcal{B}_{r_x}(\mathbf{x}) \cap \mathcal{B}_{r_y}(\mathbf{y})\}$, the region exclusive to both the circles centered at $\mathbf{o}$ and $\mathbf{x}$ as $\text{RoI}_{ox} = \{\mathcal{B}_{r_o}(\mathbf{o}) \cap \mathcal{B}_{r_x}(\mathbf{x})\} \setminus \text{RoI}_{oxy}$, the regions exclusive to both the

circles at $\mathbf{o}$ and $\mathbf{y}$ as $\text{RoI}_{oy} = \{\mathcal{B}_{r_o}(\mathbf{o}) \cap \mathcal{B}_{r_y}(\mathbf{y})\} \setminus \text{RoI}_{oxy}$, and the common region exclusive to circles at $\mathbf{x}$ and $\mathbf{y}$ as $\text{RoI}_{xy} = \{\mathcal{B}_{r_x}(\mathbf{x}) \cap \mathcal{B}_{r_y}(\mathbf{y})\} \setminus \text{RoI}_{oxy}$. Conditioned on Φ_u and $\mathbf{r}$, the first term on the right hand side of (24) can be expressed as given in (26) at the top of the next page, where $r_i = \sqrt{d_i^2 + r_o^2 - 2r_o d_i \cos(v_i)}$ for $i \in \{x, y\}$, the function $\text{AoI}_2(\cdot, \cdot, \cdot)$ is given in (14), the area of intersection of three circles is evaluated as per the procedure presented in Appendix D. Further in (26), the Step (a) follows from the fact that the regions in indicator functions are non-overlapping and

$$\begin{aligned}
& \mathbb{E}_{\{\Phi_{\mathbf{r}} \setminus \mathbf{r}\}} \left[\mathbf{1}_{|\Phi_{\mathbf{r}} \cap \mathcal{B}_{r_x}(\mathbf{x})| \leq N_s - 1} \mathbf{1}_{|\Phi_{\mathbf{r}} \cap \mathcal{B}_{r_y}(\mathbf{y})| \leq N_s - 1} \mathbf{1}_{|\Phi_{\mathbf{r}} \cap \mathcal{B}_{r_o}(\mathbf{o})| = N - 1} \right] \\
&= \mathbb{E}_{\{\Phi_{\mathbf{r}} \setminus \mathbf{r}\}} \left[\sum_{n=0}^{N-1} \mathbf{1}_{|\Phi_{\mathbf{r}} \cap \text{RoI}_{oxy}|=n} \sum_{m=0}^{N-n-1} \mathbf{1}_{|\Phi_{\mathbf{r}} \cap \text{RoI}_{ox}|=m} \sum_{p=0}^{N-n-m-1} \mathbf{1}_{|\Phi_{\mathbf{r}} \cap \text{RoI}_{oy}|=p} \mathbf{1}_{|\Phi_{\mathbf{r}} \cap \{\mathcal{B}_{r_o}(\mathbf{o}) \setminus \{\mathcal{B}_{r_x}(\mathbf{x}) \cup \mathcal{B}_{r_y}(\mathbf{y})\}\}| = N-n-m-p-1} \right. \\
&\quad \left. \sum_{q=0}^{\min\{N_s-n-m-1, N_s-n-p-1\}} \mathbf{1}_{|\Phi_{\mathbf{r}} \cap \text{RoI}_{xy}|=q} \mathbf{1}_{|\Phi_{\mathbf{r}} \cap \{\mathcal{B}_{r_x}(\mathbf{x}) \setminus \{\mathcal{B}_{r_o}(\mathbf{o}) \cup \mathcal{B}_{r_y}(\mathbf{y})\}\}| \leq N_s-n-m-q-1} \right. \\
&\quad \left. \mathbf{1}_{|\Phi_{\mathbf{r}} \cap \{\mathcal{B}_{r_y}(\mathbf{y}) \setminus \{\mathcal{B}_{r_o}(\mathbf{o}) \cup \mathcal{B}_{r_x}(\mathbf{x})\}\}| \leq N_s-n-p-q-1} \right] \\
&\stackrel{(a)}{=} \sum_{n=0}^{N-1} \text{Pos}_{\text{PMF}}(n, \lambda_{\mathbf{r}} \text{AoI}_3(r_o, r_x, r_y, v_x, v_y)) \\
&\quad \times \sum_{m=0}^{N-n-1} \text{Pos}_{\text{PMF}}(m, \lambda_{\mathbf{r}} (\text{AoI}_2(r_o, r_x, v_x) - \text{AoI}_3(r_o, r_x, r_y, v_x, v_y))) \\
&\quad \times \sum_{p=0}^{N-n-m-1} \text{Pos}_{\text{PMF}}(p, \lambda_{\mathbf{r}} (\text{AoI}_2(r_o, r_y, v_y) - \text{AoI}_3(r_o, r_x, r_y, v_x, v_y))) \\
&\quad \times \text{Pos}_{\text{PMF}}(N-n-m-p-1, \lambda_{\mathbf{r}} (\pi r_o^2 - \text{AoI}_2(r_o, r_x, v_x) - \text{AoI}_2(r_o, r_y, v_y) + \text{AoI}_3(r_o, r_x, r_y, v_x, v_y))) \\
&\quad \times \sum_{q=0}^{\min\{N_s-n-m-1, N_s-n-p-1\}} \text{Pos}_{\text{PMF}}(q, \lambda_{\mathbf{r}} (\text{AoI}_2(r_x, r_y, u_{xy}) - \text{AoI}_3(r_o, r_x, r_y, v_x, v_y))) \\
&\quad \times \text{Pos}_{\text{PMF}}(N_s-n-m-q-1, \lambda_{\mathbf{r}} (\pi r_x^2 - \text{AoI}_2(r_o, r_x, v_x) - \text{AoI}_2(r_x, r_y, u_{xy}) + \text{AoI}_3(r_o, r_x, r_y, v_x, v_y))) \\
&\quad \times \text{Pos}_{\text{PMF}}(N_s-n-p-q-1, \lambda_{\mathbf{r}} (\pi r_y^2 - \text{AoI}_2(r_o, r_y, v_y) - \text{AoI}_2(r_x, r_y, u_{xy}) + \text{AoI}_3(r_o, r_x, r_y, v_x, v_y))) \\
&= h_{\text{tag}, m_2}(r_o, d_x, d_y, v_x, v_y, N). \tag{26}
\end{aligned}$$

$\Phi_{\mathbf{r}} \setminus \mathbf{r}$ is a homogeneous PPP with density $\lambda_{\mathbf{r}}$. Similar to the derivation of the first moment, we obtain the final expression for the second moment by deconditioning over $\mathbf{r}$ and then over $\Phi_{\mathbf{u}}$ by applying Campbell's theorem.

C. Proof of Lemma 7

Much of the derivation can be done on the similar lines as that of Appendix B. Since $\Phi_{\mathbf{r}}$ is a homogeneous PPP, it is translation invariant. Hence, we assume that the typical AP is located at the origin. The mean load of the AP is

$$\begin{aligned}
\mathbb{E}[K_o] &= \mathbb{E} \left[\sum_{\mathbf{x} \in \Phi_{\mathbf{u}}} \mathbb{E} \left[\mathbf{1}_{(|\Phi_{\mathbf{r}} \cap \mathcal{B}_{\|\mathbf{x}\|}(\mathbf{x})| \leq N_s - 1)} \right] \right] \\
&= 2\pi\lambda_{\mathbf{u}} \int_{r=0}^{\infty} \sum_{l=0}^{N_s-1} \frac{(\pi\lambda_{\mathbf{r}}r^2)^l}{l!} \exp(-\pi\lambda_{\mathbf{r}}r^2) r dr \\
&\stackrel{(a)}{=} \frac{\lambda_{\mathbf{u}}}{\lambda_{\mathbf{r}}} \sum_{l=0}^{N_s-1} \int_{u=0}^{\infty} \frac{u^l}{l!} \exp(-u) du = \frac{N_s\lambda_{\mathbf{u}}}{\lambda_{\mathbf{r}}},
\end{aligned}$$

where (a) follows from replacing $u = \pi\lambda_{\mathbf{r}}r^2$. The second moment of the load can be written as $\mathbb{E}[K_o^2] =$

$$\begin{aligned}
& \mathbb{E} \left[\left| \sum_{\mathbf{x} \in \Phi_{\mathbf{u}}} \mathbf{1}_{(|\Phi_{\mathbf{r}} \cap \mathcal{B}_{\|\mathbf{x}\|}(\mathbf{x})| \leq N_s - 1)} \right|^2 \right] \\
&= \mathbb{E} \left[\sum_{\substack{\mathbf{x} \neq \mathbf{y} \\ \mathbf{x} \in \Phi_{\mathbf{u}} \\ \mathbf{y} \in \Phi_{\mathbf{u}}}} \mathbb{E}_{\Phi_{\mathbf{r}}} \left[\mathbf{1}_{|\Phi_{\mathbf{r}} \cap \mathcal{B}_{\|\mathbf{x}\|}(\mathbf{x})| \leq N_s - 1} \mathbf{1}_{|\Phi_{\mathbf{r}} \cap \mathcal{B}_{\|\mathbf{y}\|}(\mathbf{y})| \leq N_s - 1} \right] \right]
\end{aligned}$$

$$+ \underbrace{\mathbb{E} \left[\sum_{\mathbf{x} \in \Phi_{\mathbf{u}}} \mathbb{E} \left[\mathbf{1}_{|\Phi_{\mathbf{r}} \cap \mathcal{B}_{\|\mathbf{x}\|}(\mathbf{x})| \leq N_s - 1} \right] \right]}_{\mathbb{E}[K_o]}.$$

We can decompose the inner expectation

$$\begin{aligned}
& \mathbb{E}_{\Phi_{\mathbf{r}}} \left[\mathbf{1}_{|\Phi_{\mathbf{r}} \cap \mathcal{B}_{\|\mathbf{x}\|}(\mathbf{x})| \leq N_s - 1} \mathbf{1}_{|\Phi_{\mathbf{r}} \cap \mathcal{B}_{\|\mathbf{y}\|}(\mathbf{y})| \leq N_s - 1} \right] = \\
& \sum_{l=0}^{N_s-1} \mathbb{E}_{\Phi_{\mathbf{r}}} \left[\mathbf{1}_{|\Phi_{\mathbf{r}} \cap \{\mathcal{B}_{\|\mathbf{x}\|}(\mathbf{x}) \cap \mathcal{B}_{\|\mathbf{y}\|}(\mathbf{y})\}| = l} \right. \\
& \quad \times \mathbf{1}_{|\Phi_{\mathbf{r}} \cap \{\mathcal{B}_{\|\mathbf{y}\|}(\mathbf{y}) \setminus \{\mathcal{B}_{\|\mathbf{x}\|}(\mathbf{x}) \cap \mathcal{B}_{\|\mathbf{y}\|}(\mathbf{y})\}| \leq N_s - l - 1} \\
& \quad \times \mathbf{1}_{|\Phi_{\mathbf{r}} \cap \{\mathcal{B}_{\|\mathbf{x}\|}(\mathbf{x}) \setminus \{\mathcal{B}_{\|\mathbf{x}\|}(\mathbf{x}) \cap \mathcal{B}_{\|\mathbf{y}\|}(\mathbf{y})\}| \leq N_s - l - 1} \left. \right].
\end{aligned}$$

From the above expression, we obtain the expression for $h_{\text{typ}, m_2}(r_x, r_y, u)$ by taking the expectation with respect to $\Phi_{\mathbf{r}}$. Now, with the application of Campbell's formula, we get the final expression of the lemma as

$$\begin{aligned}
& \mathbb{E} \left[\sum_{\mathbf{x} \in \Phi_{\mathbf{u}}} \sum_{\substack{\mathbf{y} \in \Phi_{\mathbf{u}} \\ \mathbf{x} \neq \mathbf{y}}} h_{\text{typ}, m_2}(r_x, r_y, u) \right] \\
&= 2\pi\lambda_{\mathbf{u}}^2 \int_{r_x=0}^{\infty} \int_{r_y=0}^{\infty} \int_{u=0}^{2\pi} h_{\text{typ}, m_2}(r_x, r_y, u) du r_y dr_y r_x dr_x,
\end{aligned}$$

where r_x, r_y , and u are as depicted in Fig. 7 (Right).

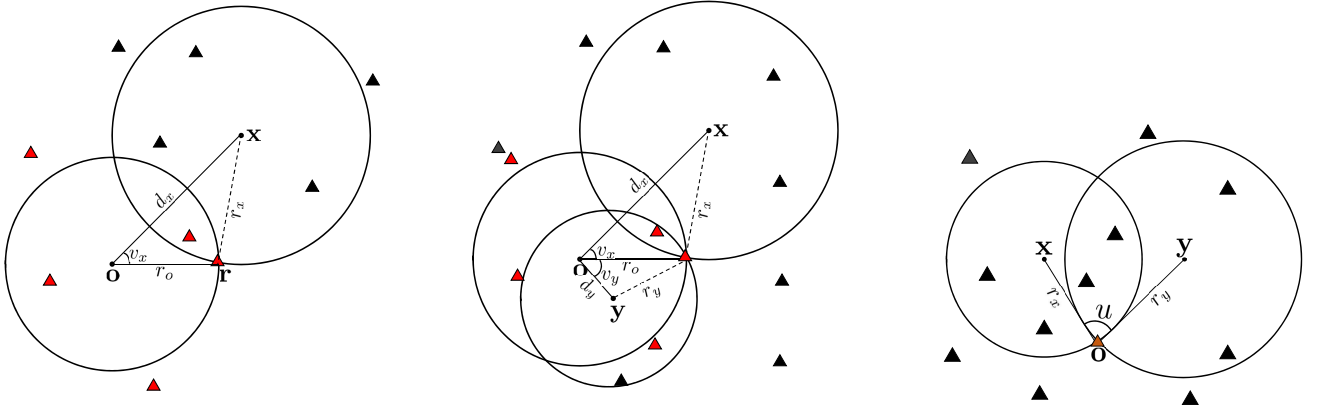


Fig. 7: (Left and center figures are for the proof in Appendix B.) The typical user is located at $\mathbf{o}$, while $\mathbf{x}, \mathbf{y} \in \Phi_u$ are random user locations. The red triangles represent the serving AP locations of the typical user. The illustration is for the third nearest serving AP for the typical user. (Right figure is for the proof in Appendix C.) $\mathbf{x}$ and $\mathbf{y}$ correspond to user locations and $\mathbf{o}$ correspond to the typical AP location.

D. Area of Intersection of Three Circles

Due to the constraint that the three circles have a common point of intersection, the common area of intersection will be either of the following three cases: (1) a point with area zero, (2) a lens, or (3) a circular triangle. All three cases are presented in Fig. 8. When the common area of intersection is a circular triangle (right most case in Fig. 8), the area is given as [54] $\text{AoI}_3(r_o, r_x, r_y, v_x, v_y) =$

$$\begin{aligned} & \frac{1}{4} \sqrt{(c_1 + c_2 + c_3)(c_2 + c_3 - c_1)(c_1 + c_3 - c_2)(c_1 + c_2 - c_3)} \\ & + r_o^2 \arcsin \frac{c_1}{2r_o} - \frac{c_1}{4} \sqrt{4r_o^2 - c_1^2} + r_y^2 \arcsin \frac{c_2}{2r_y} \\ & - \frac{c_2}{4} \sqrt{4r_y^2 - c_2^2} + r_x^2 \arcsin \frac{c_3}{2r_x} - \frac{c_3}{4} \sqrt{4r_x^2 - c_3^2}, \end{aligned}$$

where c_1, c_2, c_3 are chord lengths as denoted in the figure. Please note that the first two cases are special cases of the third case, e.g., we can get the second case by replacing $c_2 = 0$ and $c_1 = c_3$. Similarly, in the first case, $c_1 = c_2 = c_3 = 0$. Further, c_i s are functions of r_o, r_x, r_y, v_x, v_y . The procedure for determining them is outlined in [54] that is followed in this work.

E. Proof of Proposition 2

Ignoring the pilot contamination term in the expression of achievable rate in (3), we can write the rate coverage as given in (27) at the top of the next page. To proceed further, we first condition on the distance to the N_s -th serving AP $d_{N_s o}$. Conditioned on this distance, we replace $\sum_{\mathbf{r}_l \in \Phi_{r_o}^C} \beta_{lo}$ by its mean

$$\text{which is given as } \mathbb{E}_{\Phi_{r_o}^C} \left[\sum_{\mathbf{r}_l \in \Phi_{r_o}^C} \beta_{lo} \right] = 2\pi\lambda_r \int_{r=d_{N_s o}}^{\infty} l(r) r dr.$$

This result follows from the application of Campbell's theorem. Note that using the mean instead of the exact expectation has marginal impact on the accuracy of the result as $\sum_{\mathbf{r}_l \in \Phi_r} \beta_{lo}$ is dominated by contributions from the nearest N_s serving APs. Hence, we write

$$\sum_{\mathbf{r}_l \in \Phi_r} \beta_{lo} \approx \sum_{\mathbf{r}_l \in \Phi_{r_o}} \beta_{lo} + \mathbb{E}_{\Phi_{r_o}^C} \left[\sum_{\mathbf{r}_l \in \Phi_{r_o}^C} \beta_{lo} \right].$$

Next, the loads for the different serving APs are correlated. Hence, to get the exact result, we need to evaluate (27) with respect to the joint distribution of $\{K_i\}_{i=1}^{N_s}$. However, obtaining this joint distribution is not tractable. Hence, we exactly consider the load of the nearest AP and replace the load of the rest of the APs by its effective mean. For the i -th nearest AP, the effective mean is given as $\bar{K}_i = 1 + \sum_{k_i=0}^{\infty} \min\{k_i, K_{\max}\} \mathbb{P}[K_i = k_i]$, where K_i follows negative binomial distribution whose PMF is determined using the moment matching method presented in Sec. IV-A. Under the above two approximations, conditioned on the distances to the serving APs and the load of the nearest AP to the typical user, the rate coverage is given as $R_{c, \text{inf}}(T_r) =$

$$\mathbb{E}_{k_1, d_{1o}, \dots, d_{N_s o}} \left[\mathbf{1} \left(2\pi\lambda_r \int_{d_{N_s o}}^{\infty} l(r) r dr \leq h_{\text{cov}}(k_1, d_{1o}, d_{2o}, \dots, d_{N_s o}) \right) \right], \quad (28)$$

where $h_{\text{cov}}(k_1, d_{1o}, d_{2o}, \dots, d_{N_s o})$ is given by (18). Note that conditioned on $d_{N_s o}$, d_{io} for $1 \leq i \leq N_s - 1$ are i.i.d. distributed with the following PDF [55]

$$f_{D_{io}}(d_{io}) = \frac{2d_{io}}{d_{N_s o}^2}, \quad 0 \leq d_{io} \leq d_{N_s o}.$$

Further, the PDF of $D_{N_s o}$ is given as [55]

$$f_{D_{N_s o}}(d_{N_s o}) = \frac{2}{\Gamma(N_s)} (\pi\lambda_r)^{N_s} d_{N_s o}^{2N_s-1} \exp(-\pi\lambda_r d_{N_s o}^2).$$

We evaluate the expectation in (28) using the aforementioned distance distributions along with the PMF of the load K_1 associated with the nearest tagged AP.

REFERENCES

- [1] P. Parida, H. S. Dhillon, and A. F. Molisch, "Downlink performance analysis of cell-free massive MIMO with finite fronthaul capacity," in *Proc., IEEE Veh. Technology Conf. (VTC)*, 2018, pp. 1–6.
- [2] P. Parida, "Stochastic geometry perspective of massive MIMO systems," Ph.D. dissertation, Virginia Tech, 2021.
- [3] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. on Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, March 2017.

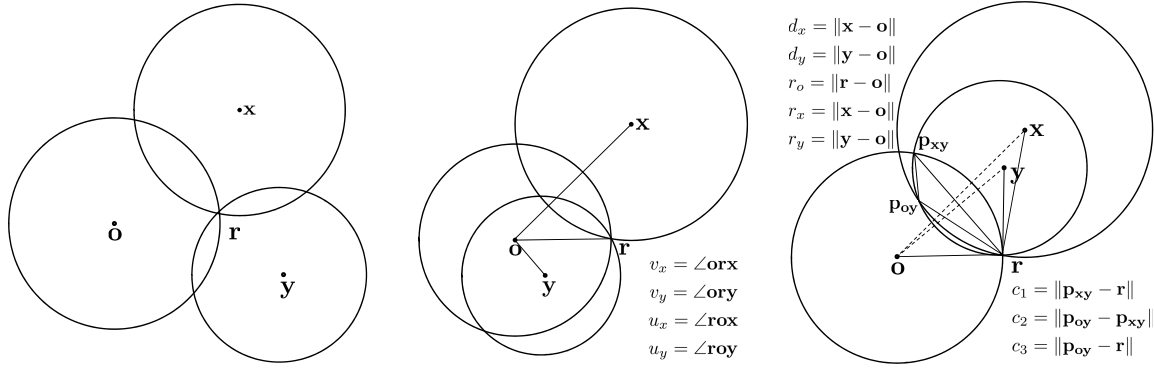


Fig. 8: Three possible configurations of the intersection of three circles with one common point of intersection r . The fourth configuration in which the smallest circle lies inside the two bigger circles has been ignored as it is a zero probability event. All relevant distances and angles have been marked in the figure.

$$R_{c,\inf}(T_r) = \mathbb{P} \left[\frac{\rho_d N_a}{2^{T_r/\xi_{\text{td}}} - 1} \left(\sum_{\mathbf{r}_l \in \Phi_{r_o}} \sqrt{\gamma_{lo}(1 - 2^{-C_\ell/k_l})/K_{\max}} \right)^2 - \rho_d N_a \sum_{\mathbf{r}_l \in \Phi_{r_o}} \gamma_{lo} 2^{-C_\ell/k_l} / K_{\max} - \rho_d \sum_{\mathbf{r}_l \in \Phi_{r_o}} \beta_{lo} - 1 \geq \rho_d \sum_{\mathbf{r}_l \in \Phi_{r_o}} \beta_{lo} \right]. \quad (27)$$

- [4] E. Nayeibi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Precoding and power optimization in cell-free massive MIMO systems," *IEEE Trans. on Wireless Commun.*, vol. 16, no. 7, pp. 4445–4459, July 2017.
- [5] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai (Shitz), "Downlink multicell processing with limited-backhaul capacity," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, no. 1, p. 840814, Jun 2009.
- [6] R. Zakhour and D. Gesbert, "Optimized data sharing in multicell MIMO with finite backhaul capacity," *IEEE Trans. on Signal Processing*, vol. 59, no. 12, pp. 6102–6111, Dec 2011.
- [7] S. H. Park, O. Simeone, O. Sahin, and S. Shamai, "Joint precoding and multivariate backhaul compression for the downlink of cloud radio access networks," *IEEE Trans. on Signal Processing*, vol. 61, no. 22, pp. 5646–5658, Nov 2013.
- [8] Q. Zhang, C. Yang, and A. F. Molisch, "Downlink base station cooperative transmission under limited-capacity backhaul," *IEEE Trans. on Wireless Commun.*, vol. 12, no. 8, pp. 3746–3759, August 2013.
- [9] S. H. Park, O. Simeone, O. Sahin, and S. S. Shitz, "Fronthaul compression for cloud radio access networks: Signal processing advances inspired by network information theory," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 69–79, Nov 2014.
- [10] T. Q. Quek, M. Peng, O. Simeone, and W. Yu, *Cloud radio access networks: Principles, technologies, and applications*. Cambridge University Press, 2017.
- [11] A. Burr, M. Bashar, and D. Maryopi, "Cooperative access networks: Optimum fronthaul quantization in distributed massive MIMO and Cloud RAN - invited paper," in *Proc., IEEE Veh. Technology Conf. (VTC)*, 2018, pp. 1–5.
- [12] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, and M. Debbah, "Cell-free massive MIMO with limited backhaul," in *Proc., IEEE Intl. Conf. on Commun. (ICC)*, 2018, pp. 1–7.
- [13] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, E. G. Larsson, and P. Xiao, "Energy efficiency of the cell-free massive MIMO uplink with optimal uniform quantization," *IEEE Trans. on Green Commun. and Networking*, vol. 3, no. 4, pp. 971–987, 2019.
- [14] G. Femenias and F. Riera-Palou, "Cell-free millimeter-wave massive MIMO systems with limited fronthaul capacity," *IEEE Access*, vol. 7, pp. 44 596–44 612, 2019.
- [15] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, M. Debbah, and P. Xiao, "Max-min rate of cell-free massive MIMO uplink with optimal uniform quantization," *IEEE Trans. on Commun.*, vol. 67, no. 10, pp. 6796–6815, 2019.
- [16] H. Masoumi and M. J. Emadi, "Performance analysis of cell-free massive MIMO system with limited fronthaul capacity and hardware impairments," *IEEE Trans. on Wireless Commun.*, vol. 19, no. 2, pp. 1038–1053, 2020.
- [17] G. Femenias and F. Riera-Palou, "Fronthaul-constrained cell-free massive MIMO with low resolution ADCs," *IEEE Access*, vol. 8, pp. 116 195–116 215, 2020.
- [18] M. Bashar, H. Q. Ngo, K. Cumanan, A. G. Burr, P. Xiao, E. Björnson, and E. G. Larsson, "Uplink spectral and energy efficiency of cell-free massive MIMO with optimal uniform quantization," *IEEE Trans. on Commun.*, vol. 69, no. 1, pp. 223–245, 2020.
- [19] S. Elhoushy and W. Hamouda, "Downlink performance of limited-fronthaul cell-free massive mimo," in *Proc., IEEE Intl. Conf. on Commun. (ICC)*. IEEE, 2021, pp. 1–6.
- [20] S. Buzzi and C. D'Andrea, "Cell-free massive MIMO: User-centric approach," *IEEE Wireless Commun. Letters*, vol. PP, no. 99, pp. 1–1, 2017.
- [21] H. Q. Ngo, L.-N. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the total energy efficiency of cell-free massive MIMO," *IEEE Trans. on Green Commun. and Networking*, vol. 2, no. 1, pp. 25–39, 2018.
- [22] G. Interdonato, E. Björnson, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Ubiquitous cell-free massive MIMO communications," *EURASIP Journal on Wireless Commun. and Networking*, vol. 2019, no. 1, pp. 1–13, 2019.
- [23] O. Y. Bursalioglu, G. Caire, R. K. Mungara, H. C. Papadopoulos, and C. Wang, "Fog massive MIMO: A user-centric seamless hot-spot architecture," *IEEE Trans. on Wireless Commun.*, vol. 18, no. 1, pp. 559–574, 2019.
- [24] E. Björnson and L. Sanguinetti, "Scalable cell-free massive MIMO systems," *IEEE Trans. on Commun.*, vol. 68, no. 7, pp. 4247–4261, 2020.
- [25] S. Buzzi, C. D'Andrea, A. Zappone, and C. D'Elia, "User-centric 5G cellular networks: Resource allocation and comparison with the cell-free massive MIMO approach," *IEEE Trans. on Wireless Commun.*, vol. 19, no. 2, pp. 1250–1264, 2020.
- [26] S. Elhoushy, M. Ibrahim, and W. Hamouda, "Cell-free massive MIMO: A survey," *IEEE Commun. Surveys and Tutorials*, vol. 24, no. 1, pp. 492–523, 2021.
- [27] Ö. T. Demir, E. Björnson, L. Sanguinetti et al., "Foundations of user-centric cell-free massive MIMO," *Foundations and Trends® in Signal Processing*, vol. 14, no. 3–4, pp. 162–472, 2021.
- [28] F. Baccelli and A. Giovanidis, "A stochastic geometry framework for analyzing pairwise-cooperative cellular networks," *IEEE Trans. on Wireless Commun.*, vol. 14, no. 2, pp. 794–808, 2014.
- [29] G. Nigam, P. Minero, and M. Haenggi, "Coordinated multipoint joint

- transmission in heterogeneous networks,” *IEEE Trans. on Commun.*, vol. 62, no. 11, pp. 4134–4146, 2014.
- [30] R. Tanbourgi, S. Singh, J. G. Andrews, and F. K. Jondral, “A tractable model for noncoherent joint-transmission base station cooperation,” *IEEE Trans. on Wireless Commun.*, vol. 13, no. 9, pp. 4959–4973, 2014.
- [31] N. Lee, D. Morales-Jimenez, A. Lozano, and R. W. Heath, “Spectral efficiency of dynamic coordinated beamforming: A stochastic geometry approach,” *IEEE Trans. on Wireless Commun.*, vol. 14, no. 1, pp. 230–241, 2014.
- [32] S. Govindasamy and I. Bergel, “Uplink performance of multi-antenna cellular networks with co-operative base stations and user-centric clustering,” *IEEE Trans. on Wireless Commun.*, vol. 17, no. 4, pp. 2703–2717, 2018.
- [33] W. Bao and B. Liang, “Optimizing cluster size through handoff analysis in user-centric cooperative wireless networks,” *IEEE Trans. on Wireless Commun.*, vol. 17, no. 2, pp. 766–778, 2018.
- [34] Z. Chen and E. Björnson, “Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry,” *IEEE Trans. on Commun.*, vol. 66, no. 11, pp. 5205–5219, 2018.
- [35] A. Papazafeiropoulos, P. Kourtessis, M. D. Renzo, S. Chatzinotas, and J. M. Senior, “Performance analysis of cell-free massive MIMO systems: A stochastic geometry approach,” *IEEE Trans. on Veh. Technology*, vol. 69, no. 4, pp. 3523–3537, 2020.
- [36] S. Kusaladharma, W.-P. Zhu, W. Ajib, and G. A. A. Baduge, “Stochastic geometry based performance characterization of SWIPT in cell-free massive MIMO,” *IEEE Trans. on Veh. Technology*, vol. 69, no. 11, pp. 13 357–13 370, 2020.
- [37] S. Elhoushy and W. Hamouda, “Limiting doppler shift effect on cell-free massive MIMO systems: A stochastic geometry approach,” *IEEE Trans. on Wireless Commun.*, vol. 20, no. 9, pp. 5656–5671, 2021.
- [38] M. Ibrahim, S. Elhoushy, and W. Hamouda, “Uplink performance of MmWave-fronthaul cell-free massive MIMO systems,” *IEEE Trans. on Veh. Technology*, vol. 71, no. 2, pp. 1536–1548, 2022.
- [39] M. Afshang and H. S. Dhillon, “Fundamentals of modeling finite wireless networks using binomial point process,” *IEEE Trans. on Wireless Commun.*, vol. 16, no. 5, pp. 3355–3370, May 2017.
- [40] P. Madhusudhanan, J. Restrepo, Y. Liu, T. Brown, and K. Baker, “Downlink performance analysis for a generalized shotgun cellular system,” *IEEE Trans. on Wireless Commun.*, vol. 13, no. 12, pp. 6684–6696, Dec 2014.
- [41] V. V. Chetlur and H. S. Dhillon, “Downlink coverage analysis for a finite 3-D wireless network of unmanned aerial vehicles,” *IEEE Trans. on Commun.*, vol. 65, no. 10, pp. 4543–4558, 2017.
- [42] T. Van Chien, E. Björnson, and E. G. Larsson, “Joint power allocation and load balancing optimization for energy-efficient cell-free massive MIMO networks,” *IEEE Trans. on Wireless Commun.*, vol. 19, no. 10, pp. 6798–6812, 2020.
- [43] J. Mecke, “On the relationship between the 0-cell and the typical cell of a stationary random tessellation,” *Pattern Recognition*, vol. 32, no. 9, pp. 1645–1648, 1999.
- [44] 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,” *J. Stat. Phys.*, vol. 181, no. 5, pp. 1678–1698, 2020.
- [45] J.-S. Ferenc and Z. Neda, “On the size distribution of Poisson Voronoi cells,” *Phys. A: Stat. Mech. Appl.*, vol. 385, no. 2, pp. 518–526, 2007.
- [46] F. Baccelli, M. Klein, M. Lebourges, and S. Zuyev, “Stochastic geometry and architecture of communication networks,” *Telecommunication Systems*, vol. 7, no. 1, pp. 209–227, 1997.
- [47] S. K. Iyer and D. Yogeshwaran, “Percolation and connectivity in AB random geometric graphs,” *Adv. Appl. Probab.*, vol. 44, no. 1, pp. 21–41, 2012.
- [48] M. D. Penrose, “Continuum AB percolation and AB random geometric graphs,” *J. Appl. Probab.*, vol. 51, no. A, pp. 333–344, 2014.
- [49] C. Stegehuis and L. Weedage, “Degree distributions in AB random geometric graphs,” *Phys. A: Stat. Mech. Appl.*, vol. 586, p. 126460, 2022.
- [50] M. Di Renzo, A. Zappone, T. T. Lam, and M. Debbah, “System-level modeling and optimization of the energy efficiency in cellular networks – a stochastic geometry framework,” *IEEE Trans. on Wireless Commun.*, vol. 17, no. 4, pp. 2539–2556, 2018.
- [51] “Further Advancements for E-UTRA Physical Layer Aspects (Release 9),” 3GPP, Tech. Rep., Mar. 2017.
- [52] G. Caire, “On the ergodic rate lower bounds with applications to massive MIMO,” *IEEE Trans. on Wireless Commun.*, vol. 17, no. 5, pp. 3258–3268, 2018.
- [53] M. Medard, “The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel,” *IEEE Trans. on Info. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.
- [54] M. P. Fewell, “Area of common overlap of three circles,” Defence Science and Technology Organisation Edinburgh (Australia), Tech. Rep., 2006.
- [55] M. Haenggi, *Stochastic Geometry for Wireless Networks*. Cambridge University Press, 2013.